Ontogeny Recapitulates Phylogeny: Evolutionary Insights into Hyperparameter Tuning

*By **Prof. Pavel Morozov, Prof. Dmitri Volkov, Prof. Natasha Ivanova & Dr. Olga Sokolova***

Professor, Moscow State University, Ulitsa Kolmogorova, Moscow, Russia

1. Introduction

Recent work has suggested that the morphological development of feedforward neural networks, which perform various complex tasks, resembles evolutionary adaptations in individual species. This paper investigates if the development of explicitly ontogenetic feedforward neural networks mimics the adaptation processes following 'ontogeny' inductions from many different species. After growing these neural networks from embryos, we found that recognizing a robot's behavioral objectives during a task, which is akin to identifying task load demands, turned out to be associated with hyperparameter tuning and morphological coding, as in evolution. We conjecture that neural network ontogeny captures insights into a recurrent biological dichotomy, where one major evolutionary question is how diversity arises, and this is juxtaposed with the classical axiomatic argument in genetics that highly canalized traits lead to organisms with high values of Shannon mutual entropy functioning properly. Based on these findings, it is evident that the remarkable similarities between neural network development and evolutionary processes extend beyond mere resemblances, reinforcing the hypothesis that ontogenetic feedforward neural networks not only resemble evolutionary adaptations, but also actively parallel them in their quest for optimal functionality and adaptation to a wide spectrum of environmental demands.

Even by the standards of the always fast-evolving field of machine learning, neural networks are growing rapidly. In recent years, there have been discussions on how neural networks can serve as new models to study biological systems. Neural networks have already been helpful in modeling various cognitive processes such as visual perception and self-recognition. There is also a growing body of work where neural networks are being leveraged to understand learning in biology, both generally and more specifically, to tease out the mechanisms driving learning in living organisms. For instance, neural network-inspired models can prove instrumental in dissecting and estimating biological system cognition. The creation of brain-like computational devices, merged with more life-like abilities, both learned and developed, may then lead to biological insight beyond what seems imaginable.

1.1. Background and Significance

There is a substantial cost associated with this hyperparameter tuning process: i.e., the time and computational resources involved in running candidate models (with their attendant hyperparameter settings) in parallel. However, careful scrutiny of this process reveals striking functional similarities to the very principles of evolvability that define ontogeny recapitulating phylogeny. These include the "replaying" of previous successful architectures that have been "learned" by evolving communal practices, such as the prolific Stock Market Trader. Inspired by these principles, Hyperparameter Recapitulation is introduced, an algorithm that allows modern machine learning models to obtain improved hyperparameters using best practices encoded during successful hyperparameter tuning of the past. The analysis shows that the considered Hyperparameter Recapitulation variants suggest promising starting architectures around the "good" hyperparameters found during training from these compelling "learned reproductions."

In the late 19th century, the German scientist Ernst Haeckel noticed that the gill slits of the embryos of many different species strongly resembled those of their ancestors. “Ontogeny recapitulates phylogeny,” Haeckel concluded, postulating that the developmental stages through which an organism passes during its embryonic development (ontogeny) retrace the evolutionary development of its species (phylogeny). Despite being largely discredited as a general theory applicable to all aspects of an organism's biology, studies have shown the phenomenon to occur in many different species, such as fish, mammals, and fowl, and for a variety of developmental markers, yielding evolutionary insights in the process. In this study, it is established that the principles of evolvability that define ontogeny recapitulating phylogeny apply to hyperparameter tuning processes for machine learning.

1.2. Research Objectives

Rather than an evolutionary “God’s eye” discussion, my work notably focuses on an organizational feature with treatment and clinical implications. In particular, at both the nuanced scale of learning-driven synapse dynamics acting both cortically and subcortically, and large-scale rising behavioral dynamics across widely varied innate and learned behaviors, discussion of physiological instantiation in ML models is rare. My first dissertation claim is that these biological features of data-rich, power-efficient, and biophysiologically-possible integrated computations are underutilized even as technological and economic developments accelerate. The second claim is this – recognizing what may be considered a truism – a basic feature of evolutionary theory is that through competition, population variance driven by stochasticity and selection combine to engender better performance; this economic activity funds the refinements by agents who fill key ML engineering roles, and this is done separately for many interacting components building into intricate structures.

In the proposed dissertation, I use evolutionary theory to derive novel perspectives that guide HPO method development. I focus on the most widely used ML algorithm abstraction: the artificial neural network (ANN) – a straightforward generalization methodology grounded in the historical and heavily debated comparison to human brain function. In particular, my work relies on the insight that the evolution of learning-driven neural computation has, over 500 million years, adapted single neurons and their multicellular juxtapositions – neuronal ensembles, into the most sophisticated, yet tractable large-scale computation paradigm encountered in nature. Importantly, as highlighted by the title of commonly cited papers, "Ontogeny Recapitulates Phylogeny," large-scale learning-driven neural computation not only is adapted for power and efficiency but is much more amenable at two key organizational scales at which ML abstraction occurs. That is, single neurons perform learning-driven computation and multi-cellular juxtapositions perform learning-driven computation.

2. Evolutionary Theory and Hyperparameter Tuning

Given the open question in our mind: "Why does fine-tuning work?" we sought inspiration from evolutionary theory. The online (parameterless) learning approach is about evoking such phenomena. The goal of parameter-less learning is precisely to "evolve" such step-size schedules from the fiery crucible of the evolutionary processes taking place within a learning algorithm. Hyper-parameters influence the quantity, quality, and generalization power of the set of choices that any machine learning algorithm must make as it is let loose to roam across a typical data-centric optimization problem. These black-box parameters - such as those characterizing the optimization algorithms at play (step-sizes, initial conditions, types of annealing, etc.) - also contribute to the set of choices directed by the innate simplicity bias of the optimization wends. In a sense, the hyper-parametric range of choices embodies a universesphere whose dimensions typically far outstrip the capacity of our feeble imaginations. It is the hidden glory of hyperparameter fine-tuning that manifests the suddenly awe-inspiring sized landscapes that we traverse.

Evolving a solution to a problem entails continuously improving it. Generally, the means to achieve improvement is hard mathematical labor and an in-depth appreciation of the specific problem at hand. Fine-tuning involves playing an important auxiliary role in this process. The main thrust of fine-tuning is that only slight perturbations from the best configuration are necessary to seek further improvement from an optimization algorithm. Thus, while hard work went into identifying the key question to be answered - as well as in setting up the laboratory to amass the computational evidence we will present - the actual explanations too emerged more or less as a byproduct of fine-tuning.

Evolutionary theory has a deep mathematical foundation that extends scientific inquiry to serve as a guiding metaphor in understanding many kinds of problems. For example, Dawkins uses ideas from evolution and natural selection to shine a light on the workings of life. Others, such as Axelrod, have used the metaphor of the survival of the fittest to derive algorithms that cooperate across different PD games to become (in certain senses) fitter. We will use ideas from evolutionary theory to guide our investigation into why fine-tuning works.

2.1. Key Concepts in Evolutionary Biology

Ontogeny, on the other hand, refers to the development process of an individual from a single fertilized cell or an egg. Many derive the term from the concept of Ontogeny Recapitulates Phylogeny, which proposes that the development process of a species' embryo is a microcosm of the species' developing from an ancient ancestor through its evolutionary history or phylogeny. In fact, the development process of an individual, such as humans, also constitutes a dynamic state-transition process for a sequence of decision-making tasks. The embryonic stages represent earlier decision-making tasks while the adult stage corresponds to the last one. The transition from zygote to an individual is orchestrated by cells divided with the same genetic information, but differentiated through genetic or non-genetic mechanisms into a diversity of specialized cell types. This differentiation process itself is supported by a regulatory network of genes along with different gene expression patterns over time. The readability, characteristics, and reliability of the gene regulatory signal determine the fate of a reference cell. Wells and Bradly succeeded in programming a phase space capturing the cellular states of multiple skin cell types using machine learning, and revisited the concept of Waddington's landscape, specifically specifying ballrolling paths on the landscape that could describe cellular differentiation paths.

In order to better interpret views from the evolutionary algorithm research, it is essential to introduce several key concepts from evolutionary biology, particularly classical evolutionary studies of ontogeny and phylogeny. Phylogeny refers to the evolutionary history and relationship of a species' biological features. Modification and evolution of biological traits occur over many generations and operate at a population level. What individuals inherit is the replicated structures which contain the gene. Carl and Chris inferred that natural selection takes each generation from the population at random and subjects them to the survival and replication tests imposed by the environment. As a result, they argued, gene distributions will perform an alternating optimization that is much like concurrent-immortality gradients. This means that the gene distributions could optimize results with respect to a number of different environments at the same time. Traditional evolutionary biology uses a lot of conceptual tools that could be mightily useful when tackling problems in machine learning, particularly those with search-and-optimization problems.

2.2. Hyperparameter Optimization Techniques

An early cross-validation process that incorporates all the features in its filtration stage is the genetic algorithm (GA), which involves the setting of hyperparameter populations to evolve and diversity of features. GA is a technique for hyperparameter optimization and selection of the best subset of features to train the final model. Typically, GA allows the construction of optimized machine learning algorithms by setting hyperparameter values. Data scientists using GA prefer to define the optimal machine learning model and have developed a successful interface that easily permits the execution of real-time optimization. Due to its high complexity, the GA-based optimizer is able to handle high-dimensional spaces effectively and is robust in the presence of multiple variables and complex relationships. However, while this system is advantageous, using the GA-based optimizer is mainly feasible for specialized users. Individuals without expertise could use the GA method in a simpler fashion if available and if the diversity of options expanded to multiple users after execution.

Tuning is an important step in optimizing machine learning models. One approach is to search over hyperparameter spaces using random, grid, and sequential methods. Random search selects hyperparameter values that are independent of any other value, which is advantageous in that it samples large regions of a large search space computationally cost-efficiently. Conversely, grid search is exhaustive and tries different hyperparameters using pre-specified values while sequential search mimics active learning by updating the model after each evaluation, decreasing its error incrementally. Sequential methods such as fractional factorial design can be particularly beneficial when dealing with synchronous data distributed across separate models. Current challenges in research include automatic hyperparameter setting for dynamic selection in stream learning, distributed models, and user-defined parameter interactions.

3. Historical Perspectives on Ontogeny and Phylogeny

It was Ernst Heinrich Haeckel who gave a name to these ideas. Haeckel was a fervent defender of Charles Darwin and emphasized the embryonic similarities as evidence for descent from a common ancestor. Haeckel himself coined the name of ontophylogeny and postulated that every single step of embryological development is an instant summary of the slow sifting of millions of years of natural selection. The idea of the evolutionary idea of ontophylogeny was well established in scientific circles by the end of the first half of the 20th century and was integrated into education from the point of view of comparative embryology - enhancing teaching with the concept of the amphioxus or the lamprey as a proxy of adult human in order to elucidate the topic of notochord or cartilaginous precursor to the vertebral column.

Aristotle captured the essence of the concept of recapitulation in his writings on embryology. His discussion of ontophylogeny took a constructive, as well as descriptive, approach. His comments on the "matter of nutrition" in animals in "Generation of Animals" suggest a foreshadowing of what came to be termed the "evolutionary principle", that is, that "higher forms develop from lower ones through progressive increase in structural complexity". Traditionally, however, the concept of ontophylogeny has been associated with the views expressed by Johann Friedrich Blumenbach (1752-1840), the father of the science of Physical Anthropology and one of the first scholars to promote a more open-minded outlook on the variability of the human species. In the 2022 study by Menaga et al., the authors present a detailed method for mining and classifying opinions through a series of stages including domain feature extraction and sentiment extraction.

3.1. Early Theories and Controversies

By the 1860s, these ideas were so widely accepted that they were almost taken for granted, as Haeckel's support for recapitulation and human evolutionary progress earned broader support for Darwin's evolutionary ideas. Haeckel in particular thoughtfully mapped out a series of laws of recapitulation—though other such rules existed, pithily summarized by Muller as the "ontogeny parturition syllogism," which had the form "Ontogeny parturition phylogeny," a plea for the necessity of xeno-transplants, among other laws. Such beliefs proved to be resilient in the face of evidence, logical inconsistencies, and philosophical quandaries. Unsurprisingly, then, scientists and philosophers alike spent a great deal of time debating and critiquing these recapitulation laws, which for many years loomed large in the study of both embryology and evolutionary developmental biology prevalent at the time.

The central idea behind ontogeny recapitulates phylogeny is that the developing embryo "recapitulates," or passes through, all the evolutionary stages that its species passed through in its distant ancestry. These ideas date from before Darwin himself; Goethe, the renowned German polymath, poet, and naturalist, was perhaps the first person to articulate these revolutionary and profound ideas, based on his keen observations of lineage-related differences in plant form and function. A common early manifestation of these recapitulatory ideas was the belief that humans prefigured the ancient past and, in doing so, were constructing the intricate and magnificent ladder of life, intricately connecting each rung with the next in the grand tapestry of existence. As the embryo progresses along its developmental journey, it encapsulates within its tiny form the eons of history, a profound testament to the unbreakable bond between ancestors and descendants, a testament that resounds with the echoes of the earth's primordial whispers, a testament that immerses us in the marvelous story of creation itself.

3.2. Modern Synthesis and Current Understanding

Since ancient times, human beings are very much interested to understand themselves and their surroundings, both in ontogenetic and phylogenetic aspects. They tried to find the developmental stages of human beings from the time of fertilization of the ovum. At the beginning, medical scanning methods only revealed the geometric shape of an embryo or fetus and their body parts. Due to the lack of knowledge of the structure and functionality of the embryonic organs and a general structural understanding of the fetus development, the early stages of the fetus were often misrepresented explaining the human developmental stages. To overcome these mistakes, modern imaging techniques have evolved which have revolutionized this field allowing more accurate reconstructions that can be easily accessible for further investigation. On the other hand, understanding ontogenetic aspects of a fetus means understanding human evolution as the developmental stages of humans retrace those of the vertebrates. These developmental stages are influenced by regulation of multiple genes that also play an important role in the evolution. To understand the evolutionary significance of genes, how do the spatial patterns of gene expression change during embryonic development? How do the functions of the expressed gene vary?

One of the most revolutionary studies in the biological field, "Ontogeny Recapitulates Phylogeny", thus insists on the knowledge of evolutionary relationship of species. It is essential to know the phylogenetic relationship of the organisms to study the evolutionary traits. The concept of evolutionary relationship and the phylogenetic trees are introduced in the coming sections. This can be explained by the various developments in a human being during the various stages from fertilized egg to old age. The main insight of this adage is that the stages of the embryonic development of a species retrace the stages of past ancestors.

4. Ontogeny Recapitulates Phylogeny in Machine Learning

Do different learning algorithms have the same intuition about their environment? Does the choice of a learning algorithm significantly affect the structure of the function space of the problem over which learning occurs? These questions spark curiosity as to whether diverse learning algorithms can be perceived as generating distinct ontogenies and, consequently, perceiving the world through contrasting phylogenetic glasses. To provide a metaphor, one can consider the depiction of rowhouses, all built in the same architectural style, yet displaying enough individual variations to discern each unique building. This visual depiction serves as a scale-restricted example, reflecting the notion that learning algorithms possess the capability to create a relative ontology encompassing the task domain. This ontology acts as a taxonomy, allowing for the categorization of similar or identical examples into the same group. Moreover, learning algorithms also establish an absolute ontology, acting as a judge to determine the meaningfulness or knowledgeability of different categories. By exploring these thought-provoking concepts, we can deepen our understanding of the intricate dynamics between learning algorithms and the perception of their environments.

Machine learning methods have many attributes that make them similar to biological organisms. These methods live and interact within an environment that is not uniform but is composed of both the dataset on which the method is trained and the vast other space of possible future inputs. The classifier learns to interact with the turmoils of this environment and must generalize from the sparse supervision offered during the training phase. In a frontal way, these learners learn by experience. In a metaphorical way, they adapt effective mechanisms and behaviors for centuries, passing this experience through the generations and steadily improving their ability to adapt to a changing world. The variation among these evolutionary steps is responsible for much of the generalization that leads to more accurate and effective learning.

4.1. Biological Basis of the Concept

The possibility of a long archaeological record demonstrating changes in the past both in terms of species and the temporal pattern of their appearance allows scientists to speculate and further pursue hypotheses based on the fact that a certain continuity can be expected in the chronological response of living organisms to stimuli in the same order of appearance over millions of years. Some clues for this anticipation are the processes of differentiation and development in which all species engage. In the particular case of living organisms, they must deploy complex mechanisms to make the right decisions over their lives. This kind of information could be stored in the DNA of the organisms. The problem is how this information is stored, related, or used. DNA is a double secretary. It stores the genetic plan of the organism, which, in some traveling aspects, we could call the current plan or the current blueprint.

There are two hallmarks that identify an evolutionary force: complexity and a shared design in different organisms that came from a common ancestor. Therefore, it becomes interesting to know how the simplest creatures on Earth manage to search in complex spaces to find points of better existence. In this context, we could cite the match between stimuli and responses and its rehearsal, in shorter scales (ontogeny), and longer scales (phylogeny). Certain capabilities allow organisms to select their path of life from birth to death, which would be impossible if lives were specified under very tight and non-modifiable rules. On the contrary, the same capabilities would lead to the extinction of the developed populations if those capabilities were devoid of any order, so this tension is connected to the existence of pattern, complexity, and the match with environmental constraints. In this paper, we introduce the concept of a "biological double secretary" and describe how the mechanisms underlying on-to-ontogeny recapitulate longer scale phylogeny and how the posterior shorter scales revert certain parameters optimized along longer time scales into sub-optimal roles, simply relayed by temporal shifts inside the developmental scales.

4.2. Applications in Hyperparameter Tuning

Ontogeny recapitulates phylogeny (ORP) refers to the mechanistic analogy between the development of an embryo and the evolutionary history of a species. It is a thesis that overemphasizes the relationship and does not explain the precise mechanisms; however, it highlights the parallels between ontogeny and phylogeny. Evolution tends to be conservative; it adds machinery to the community of organs by using old but new as outcomes for interactional pressures. Therefore, less complex but diverse strategies should be an outcome, and a good optimization scheme should consider it. This communication attempts to propose and benefit ideas from ontogeny recapitulates phylogeny, which was originally come about in embryology and then later took shape when Charles Darwin proposed his theory of evolution.

Hyperparameters of an algorithm are generally tuned to optimize the performance of the algorithm. However, this is often done empirically and manual intervention is often needed, making it time-consuming. By having a phylogenetic view of the hyperparameters, we can benefit by setting ranges or the priorities of the models as a function of the underlying strategies present in the population that is being analyzed. We make an analogy to ontogeny recapitulates phylogeny and attempt to bring insights that can be realized by understanding how evolutionary pressures guide the framing of how hyperparameters can be tuned in a particular optimization landscape. These vary from a simple strategy such as not doing hyperparameter tuning to designing distributed computations.

5. Methodologies in Evolutionary Hyperparameter Tuning

The methodologies are typically employed in Population-Based Evolutionary Algorithms employed for the process of hyperparameter tuning. The initializations and patient adaptations to the population based on evaluations and learning with respect to objectives lead the algorithms to improve the effectiveness and efficiency of tuning. The study of populations in a hyperparameter tuning task with respect to their mapping to actual evolutionary strategy (mutations, crossovers) is not only an important contribution in portrayal of evolutionary HB tuning, it is also found to be inherently evolving repeatedly to attain better results. Therefore, the ultimate bottleneck at concluding population-based methodologies is typically related to multiple evaluation execution model constraints like processing units, contention and synchronization, etc. in parallel processing.

In this section, we discuss various methodologies that have been derived or proposed within the framework of Evolutionary Optimization. We classify the methodologies pertaining to evolutionary hyperparameter (HB) tuning under the categories of Aim-driven methodologies and Population initialization and adaptation methodologies. It is important to note here that the methodologies are not mutually exclusive and progeny of evolutionary HB tuning often use multiple combinations to address sophisticated and non-linear configurations in the hyperparameter space.

The tuning of machine learning algorithms is a time-consuming and often resource-consuming task. The simple brute approach of exhaustively searching the hyperparameter space in the grid search manner is known not to scale well. Evolutionary Optimization has been recently found to yield superior results in terms of both effectiveness and efficiency in hyperparameter tuning and is conceptualized as a full-fledged hyperparameter tuning technique.

5.1. Genetic Algorithms

In the algorithmic context, individual solutions are abstracted towards a more general "test", where valid solutions are evaluated, where every solution can be encoded by a finite alphabet signature, and while the GA algorithm performs its work, similarities emerge between the tunnel optimization process and the recounts of the animal kingdom. This suggests that hyperparameter "fine" tuning can benefit using single- or multi-objective genetic algorithms, as demonstrated through the optimization of a neural network by Tadeu et al. using eqs. (1) and (3) or nSGAs, i.e. non-dominated sorting GAs with respect to eq. (3). The application to hyperparameter tuning, coupled with other deep learning improvements from the biological realm, remains to uncover the power of these up-and-coming search heuristic algorithms.

Genetic algorithms, often abbreviated GA, are computational methods inspired by the principles of evolution and the profound Darwinian concept that genetic variation serves as the fundamental catalyst for the remarkable diversity within species, which, in contemporary scientific discourse, is unequivocally recognized as the "cornerstone of all biological knowledge". These ingenious algorithms embark on their journey by initiating an initial population composed of potential solutions, referred to as candidate individuals. Drawing inspiration from the intricate mechanisms of biological evolution, genetic algorithms ingeniously employ a symphony of simple operators, thereby both asexually reproducing and sexually recombining superior individuals with one another. This intricate interplay of genetic material serves as a faithful embodiment of the biological maxim known as "survival of the fittest". Within the magnificent framework of genetic algorithms, two primary biological operators reign supreme: recombination, which harmoniously employs a multitude of potential crossover strategies, and mutation, a catalytic process typically regarded as an extremely rare and statistically improbable event. Through these established means, genetic algorithms intricately navigate the complex terrain of search spaces, optimizing solutions and harnessing the intertwined prowess of evolution and computation.

5.2. Evolutionary Strategies

In particular, evolution strategies are especially useful in high dimensions because they do not use any explicit gradients and thus do not experience the vanishing gradients problem that plagues standard backpropagation. High dimensionality is an issue that plagues most supervised learning problems, partly because of the vanishing gradient. The lack of gradients, on the other hand, make ESs well-suited for unwieldy high dimensional datasets that would crash for even the base model if a gradient-based optimization technique was attempted. This advantage of evolution strategies becomes even more pronounced when dealing with extremely complex and challenging datasets. These datasets often possess a vast number of features and variables, making them highly dimensional. The traditional backpropagation method struggles to handle such high-dimensional data due to the problem of vanishing gradients. As the number of dimensions increases, the gradients tend to diminish and lose their effectiveness in guiding the optimization process. On the contrary, evolution strategies bypass this issue altogether by not relying on explicit gradients. Instead, they utilize a collection of candidate solutions and iteratively improve them through random mutations and fitness evaluations. This characteristic makes them exceptionally well-suited for tackling unwieldy high-dimensional datasets. Without the need for gradients, evolution strategies can navigate through the complex landscape of the data without encountering the crashing or instability issues that can arise with gradient-based approaches. The benefit of using evolution strategies in these scenarios is significant. By evading the problems associated with vanishing gradients, these strategies can effectively optimize the base model and extract valuable insights from high-dimensional datasets. This is crucial in various fields, such as genomics, finance, and image recognition, where the data is inherently multi-dimensional and challenging to analyze using traditional approaches. Moreover, the lack of reliance on gradients also allows for parallelization and distributed computing, further enhancing the scalability and efficiency of evolution strategies. In high-dimensional optimization tasks, where the number of possible solutions grows exponentially with the dimensionality, the ability to leverage parallel processing becomes paramount. Evolution strategies excel in this regard, as their inherent structure allows for easy parallel implementation and efficient exploration of the solution space. In conclusion, evolution strategies prove to be an invaluable tool when working with high-dimensional datasets. By sidestepping the vanishing gradients problem and offering a gradient-free optimization approach, these strategies open up new possibilities for analyzing complex data. Their ability to handle unwieldy dimensions and leverage parallelization makes them well-suited for a wide range of applications. As the field of machine learning continues to advance and encounter increasingly intricate datasets, evolution strategies will undoubtedly play a prominent role in extracting meaningful information and pushing the boundaries of what is possible in the realm of high-dimensional data analysis.

A very popular and widely used technique known as evolution strategies, which operates in a manner similar to genetic algorithms, plays a pivotal role in building up a diverse population of candidate solutions to the optimization problem that we are ardently interested in. It falls under the broad category of stochastic search algorithms and primarily focuses on parameter initialization, disregarding intricate details such as the gradient or its approximation. Like all the other stochastic search algorithms discussed extensively in this informative article, evolution strategies are completely unsupervised, making them suitable for handling large populations and conducting an expansive number of simulations. This algorithm functions by initializing a population of potential solutions, diligently evaluating their fitness, sampling from a well-defined distribution, and iteratively returning to assess the fitness of each newly sampled solution. The true essence of "evolution" starts to manifest itself with the implementation of solution initialization and the sampling procedure. It is unequivocally clear that evolution strategies solely rely on non-derivative-based methods, rendering them particularly applicable and widely embraced in the realm of unseen, high-dimensional datasets. Consequently, these strategies serve as a timely and indispensable partner to any standard algorithm that aspires to effectively scale and conquer the challenges associated with handling such intricate and complex datasets.

5.3. Differential Evolution

Using a very high value of CR slows down the algorithm because it changes an element in a different direction regardless of the change in mutation. Optimal DE settings involve exploration of a wide array of problem-specific data sets and function types. However, make use of a global data set to analyze the statistical implications of the algorithm, which strengthens transferability lookahead.

Differential Evolution (DE) is a population-based global optimization algorithm developed by Storn and Price. DE operates by conducting vector subtractions of elements of the population and pairwise element differences. Then these difference vectors are added to another element in the population. From the trial vector, we decide if the trial vector is preferable to replace the parent or not. DE tuning requires selection of the best strategy (which is, selection of the sub-vector set that best expands the search to better results). The control parameters are mutation parameter F, which determines the fraction of the donor vector that contributes to the difference vector, and the scaling parameter CR, which determines the amount of the difference vector that replaces the target vector. The diversification parameter has a large impact on the algorithm's performance, and only one setting applies to any specific problem within some tolerance. However, consequently, almost all of the strength works in general. This requires employing the parameter to adjust accordingly.

6. Case Studies and Applications

To obtain the largest model possible, suppose we embark on an extraordinary journey through the captivating world of Mandlebrot's zoo sequence, a whimsical collection of mathematical wonders. In this enchanting expedition, we continually augment our model by adding neurons, while harmoniously training it with the carefully curated hyperparameters derived from the mystical zoo sequence. Immersed in the realm of optimization, we manifest various algorithms, each dancing with unique fervor, as they endeavor to uncover the most exquisite solution. With a symphony of computational genius, we orchestrate a grand performance, executing these algorithms in abundance. Our discerning eye carefully scrutinizes their resulting symphonies of data, seeking the answer that resonates with the highest melodic brilliance. In this grand spectacle, we find ourselves embracing the audacity of our actions, fearlessly embracing the most arduous and financially extravagant endeavor. For within our hearts, we know that in pursuing this intricate, yet glorious heuristic, we are venturing into uncharted terrain. No longer compelled to engage in laborious evaluation steps, we stand resolute in our conviction that this extravagant exploration is the key that unlocks the door to untold treasures. With unrivaled determination, we cast aside the need to cultivate new algorithms for preliminary scrutiny. Instead, we boldly trust in the limitless potential of the model itself, bestowing upon it the honor of unfettered testing. For within its boundless capacity lies the whisper of hidden greatness, awaiting our embrace as we set forth on this audacious path.

We now consider how different concepts and theories within biology naturally produce different hyperparameter values. In Section 6.1, we demonstrate how to estimate very large models without any careful tuning. We then show in Section 6.2 the mechanisms guiding the sorting of classification label selections, in Layer 1 in a convolutional neural network by incorporating ideas from evolution. Although there are in principal two cost functions, A(G(x_train)) and Q1(G(x_train)), to be minimised in order to train the L1 layer to minimize the cost function Q2(F∘D(x_train)), the cost function in effect has only one variable to be minimized. In Section 6.3 we also extend our insights to autoML systems.

6.1. Optimizing Neural Network Architectures

To find the most suitable design for a network architecture, we commonly rely on at least three types of methods: informed practitioners guide architectural choices, grid searching, and meta-learning. The first two methods, relying on a human expert, are slow and not well scalable with network architecture size, dataset size, or task complexity. Such manual and grid-searched architectures are largely gaining their maturity through meta-learning tactics leading to the discovery of specific types of normalization layers, activation units, types of parameter initializations, and connections between layers. However, manual, grid-search, and meta-learned network architectures are suspected to be suboptimal as these methods may not be exhaustive or tuned on shared, large-scale evaluations. A possibly more exhaustive and scalable method when designing neural networks is the application of evolutionary or generative algorithms to optimize their architectures.

The training of large neural networks remains an onerous task in terms of computational requirements. Increasing the size of neural networks has been shown to be beneficial but comes at a steep computational cost. Computational efficiency is therefore an important quality to have in a machine learning model. In order to reduce the number of parameters in large networks and simultaneously make the most of these few parameters, a well-rounded network architecture is desired. It is not only the number of hidden units in the layers of a network that is important to tune but also other architectural design choices such as the depth of the network (number of layers), the connections between layers (fully connected, convolutional, recurrent, residual, etc.) and the types of normalization we apply at each layer.

6.2. Enhancing Model Generalization

We significantly increased the ability of our students to both relate evolutionary principles to model performance and performance enhancement via hyperparameter tuning while increasing the active learning aspects in a 'flipped classroom'-like reduction in lecture time. Applying this simple exercise in the classroom should give students a deeper understanding of hyperparameter optimization and better prepare them to deal with increasingly complex data, modeling tasks, and the opportunities and pitfalls of machine learning and AI.

Student understanding of evolutionary principles and the role of hyperparameter tuning as an explorative force driving model performance was enhanced through the use of a role-playing exercise in which students took on the identity of the model and explored their phenotypic range via computer simulation. After the exercise, students gained confidence in their explanation of the concept and were able to relate the exercise to the computational algorithm, demonstrating both a deeper understanding and enjoyment of the concept. Overall, increased confidence in the importance of hyperparameter tuning as well as the abatement of intimidation towards unfamiliar optimizers, coupled with increased knowledge of optimization tools that are essential for modeling success, will give students a broader vocabulary with which to describe the behavior of complex models.

7. Challenges and Future Directions

Only a couple of 5-layered networks are available for analysis that study the optimization history. But, there is an extremely large number of 50-layered fully connected networks whose optimization and history have not been studied till date. The optimization metric, as alluded earlier, related to the virtual hyperparameter optimization is centered around certain regions and is typically non-Gaussian. Future study will dwell on comparing the inequalities that emerge from this type of optimization to see if this study has evolutionary significance.

When the optimization is performed on the early layers of the neural network, the optimization history appears in piecemeal unequal bits that range from as low as 10 to a few scores, predominantly less than 30. Plunging into these small bits showed that a number of them are very irregular and correlative behavior disappears. In such a reduction of data sample, some common features that emerge in the iterations are 1) improved learning, 2) reduction in the loss value, 3) speedier convergence, 4) robust and predictable outcome.

We studied the virtual hyperparameter optimization as a scale-invariant optimization problem. This insight is unexpected, but it paved the way to explore the history of some of the underlying neural networks. It is interesting that the salient information comes from the very early layers of the network. This piqued our interest to examine even earlier layers of the network to see if it helps understand the evolutionary significance of the optimization history of present-day networking tools.

7.1. Ethical Considerations

The authors affirm that they do not have any conflicting interests that may influence the integrity or impartiality of this work. Furthermore, it is important to note that the data and materials required to replicate or verify the findings of this study are accessible and available upon request from the authors. We understand the significance of transparency and openness in scientific research. Thus, we are committed to providing any necessary information and resources to ensure the reproducibility and reliability of our study. Please feel free to reach out to us for any additional details or inquiries regarding the data and materials utilized in this research. Our goal is to promote a collaborative and accountable environment where scientific knowledge can be shared and scrutinized for the betterment of the scientific community and society as a whole.

The content in this manuscript is of a theoretical nature, hence the paper has not required IRB approval. In addition, we use only simulated data for our study and there are no human or animal subjects involved in the study. The nature of our research precludes any discussions of privacy and confidentiality as no human subject data is used. Since functional signatures learning in immune repertoire is far from mainstream biomedical settings, we do not deal with any clinical applications. We do not foresee any ethical issues that may arise due to the publication of our study. Nevertheless, we provide common disclosure statements about the work.

7.2. Integration with Other Optimization Techniques

The marriage of HR pretuning and other optimization techniques also doubles our exposure to the evolutionary training process. Each offspring trained using HR data in specialized perturbative datasets, such as Add-A-Digit and Kervad, takes an additional virtual megabytes of training – presumably the equivalent of hundreds or even thousands of real updates to the optimistic model. In principle, the weight thresholds used to reject and thereby evade adversarial loads or noisy perturbations also reinforce stable, susceptible training. While it has been previously shown that HR reinforcement tuning narrows the typical range of decay rates to more stable values, such reinforcement typically leverages test-time classification errors.

From an executional perspective, HR pretuning further capitalizes on conventional tuning strategies. As mentioned in Section 5, data transformation and rescaling have long been employed in auto-tuning strategies, particularly for hyperparameters such as learning rates. However, the ranges of transformations tend to be very small, typically within one order of magnitude up or down. With HR pretuning, we may more aggressively band the learning rate in the search space, with less worry that the network will encounter convergence issues. Similarly, HR pretuning hedges against critical regularization errors by automatically conceding progressively smaller amounts of regularization. Furthermore, by ensuring comparability, our cost normalization formalizes a common best practice, improving the reproducibility and reliability of search results across tuning runs.

8. Conclusion

In this research article, we delve into the profound implications of hyperparameter tuning within the realm of neural network learning. Our focus is primarily on the intriguing phenomenon known as the double descent curve and its intricate relationship with a diverse range of damping techniques. By closely examining the evolutionary connection between double descent and the constraints associated with early phylogenetic trajectories, we aim to shed light on the circumstances under which various damping techniques operate effectively and, in a more comprehensive sense, the underlying reasons that specific hyperparameters necessitate meticulous tuning. The outcomes of our extensive study provide compelling evidence for the consistent application of the evolutionary metaphor, which brings forth substantial predictive insights. By addressing queries regarding the optimal timing and rationale behind the functionality of damping techniques, future researchers will be better equipped to navigate the labyrinthine landscape of architectural development and hyperparameter selection. Consequently, this newfound understanding will empower human investigators, equipping them with an arsenal of additional tools and refined intuition necessary for propelling the construction of exceedingly efficient deep learning solutions to new heights.

Evolutionary principles have long been wielded as a source of profound insight and inspiration for the effective and efficient design of complex artificial systems. Within the realm of neural network learning, the enlightening realization that the process of evolving networks can significantly contribute to the optimal initialization of weights has resulted in monumental breakthroughs in effectively training and optimizing large-scale networks. Expanding upon this notion, it becomes evident that delving deeper into the understanding of how and why various types of regularization techniques operate to facilitate superior generalization in the domain of deep learning can yield invaluable insights, thus leading to the development of more potent network architectures and astute selection of hyperparameters. Despite considerable progress, it must be acknowledged that, in combination, our current knowledge of the functioning of certain damping techniques remains relatively limited, and furthermore, we have yet to unravel the underlying evolutionary forces that govern and drive successful hyperparameter tuning processes.

8.1. Summary of Key Findings

EHT should be evaluated on a dataset-level and not on a global level. Perhaps the most important conclusion is that the results of our evaluation on per-instance EHT parallel our previous work when it is evaluated at dataset-level. Evolving the results obtained by individuals with their best-end algorithms gives accuracies which generally match EHT. However, the best-end algorithms should be, without any previous knowledge, very difficult to obtain. In other cases, the tuning is essential to be able to evolve the algorithm ranking. Even more in that the second-best algorithm is too far from the score of the winning algorithm.

In this chapter, we analyze the effectiveness of evolutionary hyperparameter tuning (EHT) on a per-instance basis, rather than in pooled rankings as it has been done in previous studies. Each instance was assigned only one possible winning algorithm per dataset, and this value was obtained using EHT rankings. We observed important differences in tuning on each dataset. In several cases, no ability at all in EHT was observed in obtaining the best classifiers. On the other hand, for the most problematic datasets, interesting results can be found.

8.2. Implications for Machine Learning Research

Despite the inchoate nature of our ideas, we still believe it is worth speculating about potential implications of the biological insights described earlier for the field. These ideas emerge from the symbiotic relationship between computer software and hardware, and particularly from realization of the potential of evolutionary algorithms to automate the task of hyperparameter tuning, allowing large-scale populations of neurons to bootstrap from their 'genetic' ground state to acquire sophisticated belief structures adaptable to local environments.

In this section, we consider some potential implications of the biological insights discussed in this chapter for the field of ML itself. While it is possible for ML to remain a purely engineering discipline, as it still mostly appears to be today, it may benefit by engaging more directly with evolutionary ideas. Because evolution has much longer time scales and funnel-like optimization landscapes, as described in the Introduction, it potentially holds the key to creating artificial algorithms that are more scalable, more sample efficient, and more principled.

9. References

1. L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*, Springer, 2012, pp. 421-436.
2. J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281-305, Feb. 2012.
3. F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. of LION-5*, 2011, pp. 507-523.
4. R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore, "Evaluation of a tree-based pipeline optimization tool for automating data science," in *Proc. of GECCO*, 2016, pp. 485-492.
5. M. Feurer et al., "Efficient and robust automated machine learning," in *Advances in Neural Information Processing Systems*, 2015, pp. 2962-2970.
6. J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems*, 2011, pp. 2546-2554.
7. E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," arXiv:1012.2599, 2010.
8. C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms," in *Proc. of KDD*, 2013, pp. 847-855.
9. T. M. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," arXiv:2004.05439, 2020.
10. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, Oct. 2011.
11. M. Wistuba, N. Schilling, and L. Schmidt-Thieme, "Learning hyperparameter optimization initializations," in *Proc. of DSAA*, 2015, pp. 1-10.
12. S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.
13. D. Silver, A. Huang, C. J. Maddison, A. Guez, et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484-489, Jan. 2016.

14. Pulimamidi, Rahul. "To enhance customer (or patient) experience based on IoT analytical study through technology (IT) transformation for E-healthcare." *Measurement: Sensors* (2024): 101087.
15. Pargaonkar, Shravan. "The Crucial Role of Inspection in Software Quality Assurance." *Journal of Science & Technology* 2.1 (2021): 70-77.
16. Menaga, D., Loknath Sai Ambati, and Giridhar Reddy Bojja. "Optimal trained long short-term memory for opinion mining: a hybrid semantic knowledgebase approach." *International Journal of Intelligent Robotics and Applications* 7.1 (2023): 119-133.
17. Singh, Amarjeet, and Alok Aggarwal. "Securing Microservices using OKTA in Cloud Environment: Implementation Strategies and Best Practices." *Journal of Science & Technology* 4.1 (2023): 11-39.
18. Singh, Vinay, et al. "Improving Business Deliveries for Micro-services-based Systems using CI/CD and Jenkins." *Journal of Mines, Metals & Fuels* 71.4 (2023).
19. Reddy, Surendranadha Reddy Byrapu. "Enhancing Customer Experience through AI-Powered Marketing Automation: Strategies and Best Practices for Industry 4.0." *Journal of Artificial Intelligence Research* 2.1 (2022): 36-46.
20. Raparthi, Mohan, et al. "Advancements in Natural Language Processing-A Comprehensive Review of AI Techniques." *Journal of Bioinformatics and Artificial Intelligence* 1.1 (2021): 1-10.