2019

Variational Autoencoder Based Estimation Of Distribution Algorithms And Applications To Individual Based Ecosystem Modeling Using EcoSim

Sourodeep Bhattacharjee
University of Windsor

VARIATIONAL AUTOENCODER BASED ESTIMATION OF DISTRIBUTION ALGORITHMS AND APPLICATIONS TO INDIVIDUAL BASED ECOSYSTEM MODELING USING ECOSIM

by

Sourodeep Bhattacharjee

A Dissertation
Submitted to the Faculty of Graduate Studies
through the School of Computer Science
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

VARIATIONAL AUTOENCODER BASED ESTIMATION OF DISTRIBUTION ALGORITHMS AND APPLICATIONS TO INDIVIDUAL BASED ECOSYSTEM MODELING USING ECOSIM

by
Souordeep Bhattacharjee

APPROVED BY:

F. Guichard, External Examiner
McGill University

K. Drouillard
Great Lakes Institute for Environmental Research

B. Boufama
School of Computer Science

I. Ahmad
School of Computer Science

R. Gras, Advisor
School of Computer Science

June 07, 2019

Declaration of Co-Authorship / Previous Publication

I. Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

I have used the tool EcoSim in the thesis, which was originally developed by Robin Gras et al., 2009 (Gras R, Devaurs D, Wozniak A, Aspinall A. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. Artificial life. 2009 Oct;15(4):423-63).

Chapter 4 and Chapter 5 of the thesis was co-authored with Brian MacPherson under the supervision of Dr. Robin Gras. In all cases, the key ideas, primary contributions, experimental designs, data analysis, some of the interpretation, and most of the writing were performed by the author, and the contribution of Brian MacPherson was primarily through the provision of forming the hypothesis to test along with Sourodeep Bhattacharjee, extending the interpretation of the results in the domains of Ecology and Biology and providing comparative analysis of the results received with related research.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis,

and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Previous Publication

This thesis includes three original papers that have been previously published/-submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication title/full citation	Publication status*
3	Estimation of Distribution using Population Queue based Variational Autoencoders / S. Bhattacharjee, R. Gras. Estimation of Distribution using Population Queue based Variational Autoencoders. 2019 IEEE CONGRESS ON Evolutionary Computation	In Press
4	A Comparison of Sexual Selection versus Random Selection With Respect to Extinction and Speciation Rates using Individual Based Modeling and Machine Learning / Bhattacharjee, Sourodeep, Brian MacPherson, and Robin Gras. A comparison of sexual selection versus random selection with respect to extinction and speciation rates using individual based modeling and machine learning. Ecological Complexity 36 (2018): 126-137.	Published
5	Animal communication of fear and safety related to foraging behavior and fitness: an individual-based modeling approach / Animal communication of fear and safety related to foraging behavior and fitness: an individual-based modeling approach Bhattacharjee, Sourodeep, Brian MacPherson, and Robin Gras. Ecological Informatics (2019)	Submitted

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at

the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Individual based modeling provides a bottom up approach wherein interactions give rise to high-level phenomena in patterns equivalent to those found in nature. This method generates an immense amount of data through artificial simulation and can be made tractable by machine learning where multidimensional data is optimized and transformed. Using individual based modeling platform known as EcoSim, we modeled the abilities of elitist sexual selection and communication of fear. Data received from these experiments was reduced in dimension through use of a novel algorithm proposed by us: Variational Autoencoder based Estimation of Distribution Algorithms with Population Queue and Adaptive Variance Scaling (VAE-EDA-Q AVS).

We constructed a novel Estimation of Distribution Algorithm (EDA) by extending generative models known as variational autoencoders (VAE). VAE-EDA-Q, proposed by us, smooths the data generation process using an iteratively updated queue (Q) of populations. Adaptive Variance Scaling (AVS) dynamically updates the variance at which models are sampled based on fitness. The combination of VAE-EDA-Q with AVS demonstrates high computational efficiency and requires few fitness evaluations. We extended VAE-EDA-Q AVS to act as a feature reducing wrapper method in conjunction with C4.5 Decision trees to reduce the dimensionality of data.

The relationship between sexual selection, random selection, and speciation is a contested topic. Supporting evidence suggests sexual selection to drive speciation.

Opposing evidence contends either a negative or absence of correlation to exist. We utilized EcoSim to model elitist and random mate selection. Our results demonstrated a significantly lower speciation rate, a significantly lower extinction rate, and a significantly higher turnover rate for sexual selection groups. Species diversification was found to display no significant difference.

The relationship between communication and foraging behavior similarly features opposing hypotheses in claim of both increases and decreases of foraging behavior in response to alarm communication. Through modeling with EcoSim, we found alarm communication to decrease foraging activity in most cases, yet gradually increase foraging activity in some other cases. Furthermore, we found both outcomes resulting from alarm communication to increase fitness as compared to non-communication.

Dedication

To my wife Debarati,

without her endless support and patience this would have been impossible.

Acknowledgements

I would first like to express my sincere gratitude and appreciation to Dr. Robin Gras, my supervisor, for his support and guidance throughout the course of this research. I would also like to thank the other members of my thesis committee for their valuable time and helpful suggestions.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The standard process of study within behavioral ecology is to first observe and interpret animal behavior and then form subsequent testable hypotheses [1]. Behavioral ecology, as a subset of all biological disciplines, additionally pays distinct consideration to interactions between organisms and the environment [2]. As animal behaviors and ecosystems, in addition to their inter-lying interactions, have the propensity to become exponentially complex, a bottom-up approach of study based on individual traits and behaviors is essential [3]. Individual based modeling facilitates the intricate study of discrete organisms as well as their involvement with other organisms and environmental conditions, such as food and predation. Through the creation of an artificial ecosystem, an entire set of interactions gives rise to high-level phenomena that emerge generating the same patterns observed in nature. Speciation, extinction, population migration, and the shape of spatial distribution of individuals are all observable events within artificial ecosystems [4].

Providing a foundation suitable to handling the complexities involved in the characterization of major ecosystem dynamics, in addition to developing predictions regarding behavior [3], is ecological modeling – an emerging hybridization of theoretical

ecology, mathematics, and computer science [5]. Within ecological modeling, artificial ecosystems offer benefits distinct from field studies as they illuminate large-scale views of evolution of systems; this enables a deep understanding of theoretical concepts concerning evolutionary process, speciation, and extinction [6]. In this respect, ecosystem simulations could also provide a vast amount of data related to each particular individual. Such insights may be difficult to measure or even infeasible in nature and, thus, the generation of this raw multidimensional data can be invaluable for use in analysis. Data analysis involves collecting, processing, cleaning, transforming, and modeling data in order to produce useful knowledge from which conclusions can be drawn [7], [8].

Machine learning, an approach of data analysis, could be used to extract useful knowledge from large datasets and propose insights. By learning from raw input data, machine learning could also aid in the decision making process [9]. Machine learning methods include regression, classification, feature selection, and rule extraction. Presence of irrelevant features (containing irrelevant, superfluous, and redundant information) affects the reliability and interpretability of knowledge processed by machine learning methods [10]. A class of algorithms known as feature selection aims to ameliorate this issue by identifying and removing datasets of irrelevant features prior to the construction of the predictive model.

Wrapper methods are a type of feature selection algorithm with a key distinction from other feature selection approaches – the feature selection step is concomitantly integrated with the model building step [11]. Some wrapper methods can be defined as a combinatorial optimization search procedure in the complete feature space producing features conforming to the predictive model. The produced subsets of features are then utilized in the evaluation phase to train and test specific classification models. Subsequently, a decision model is generated and the resulting accuracy of

this model determines the suitability of the feature subset. Therefore, the search is wrapped around the classification model in the search of the feature space for the optimal subset.

Estimation of distribution algorithms (EDA) are metaheuristics that assist in combinatorial optimization by modeling and maintaining a population of favorable candidate solutions [12–14]. The candidate solutions with the best fitness are selected and utilized to generate a probabilistic generative model. This model produces the next generation of promising candidates, and the process repeats until the optimal solution is determined or fitness of the current best solution is deemed as sufficient. Fitness is defined as problem specific criteria applied to the candidate as a whole with a foundation based upon values of each constituent feature in the candidate solution. Thus, the fitness function can be defined, for example, by the accuracy of the predictive model. Nevertheless, the precise combination of assignments to variables in the selection of higher fitness can be difficult to determine. EDAs have demonstrated favorable outcomes when applied as wrapper methods for classification [15] and [16]. Authors in [16] applied Constrained Model Search Space Bayesian Optimization Algorithm (CMSS-BOA) [17] as a wrapper method with C4.5 decision trees, and the resulting fitness was evaluated based upon the predictive accuracy of the classifier. We have proposed a new variant of EDA - Variational Autoencoder EDA with Population Queue (VAE-EDA-Q) which is an approach that utilizes variational autoencoders to model the EDA population and has demonstrated efficacy in exploring latent continuous search spaces [18], [19]. VAE-EDA-Q was utilized as a wrapper method to discover relevant features in multidimensional data within two areas of research using EcoSim [20], which is a tri-trophic ecosystem simulator. Applications of individual based modeling in conjunction with machine learning provide the opportunity to shed some light on real world debates in the area of biology and ecology.

Within biological literature, the distinction between sexual selection and panmixia (random mating) is undisputed, yet the implications of sexual selection remain a subject of discussion. Demonstration of the presence of panmixia is presented by a number of studies, including [21]. However, the relationship between random mating and speciation and extinction rates is not universally understood based upon existing empirical study alone. Further, empirically based dialogue within the concept of sexual selection, specifically, also lacks concordance. Authors in [22] propose sexual selection to involve variance due to mating success, and natural selection to involve variance with respect to other aspects of fitness. And while the authors of [22] define sexual selection as intra-specific reproductive competition, they also admit it to be a poorly understood concept. Within sexual selection, the concept known as the good genes hypothesis is another fundament that remains open to debate. This concept is based upon the assumption that females who select males with phenotypic traits presumed to manifest good genes will produce fit offspring [23]. While the meta-analysis performed by [23] revealed a correlation between male secondary traits that attract females during mating and offspring survival, authors in [24] determined the role of male secondary traits to be minor in the selection for good genes within Pronghorn.

By departing from labor intensive field study and instead focusing on simulative study, we have propose to examine sexual selection as related to speciation and extinction rates using a tri-trophic ecosystem simulator named EcoSim [20], [25]. In our control group of runs we allowed female prey to select their mates randomly. While in our test runs the females were given the faculty to pick mating partners with highest strength. We have chosen to employ VAE-EDA-Q in combination with decision trees to build predictive models for the examination of the aspects of rates of speciation and rates of extinction population wide. By creating these models, we propose to gain a deeper understanding of the longstanding literary debates described above.

Furthermore, the data garnered from our study can be translated to valuable insight reaching far beyond the original empirical queries.

The subject of behavioral influence of animal communication displays discordance similar to that of sexual selection. Evidence exists to support that communication influences behavior, as described by [26] in his study of the influence of honey bee waggle dances upon foraging behavior. However disagreement exists over the direction to which the influence is an effector (increase versus decrease). Within the scope of foraging behavior, there is a high level of empirical corroboration for predator alarm cues and presence of predation decreasing foraging behavior in prey, with specific cases presented in the studies of hard clams, coral reef fish, termites, and crabs [27], [28], [29], [30]. However, ample opposition exists in support of a gradual, long run increase in foraging behavior due to sustained predation and communication of alarm cues – this is known as the predation risk allocation hypothesis. This view employs a cost-benefit analysis to reason that the cost of vulnerability to predation is outweighed by the benefits of acquiring food necessary for survival [31]. Additional corroboration of this theory is presented by a number of studies including [32], [33], [34]. Despite this evidence, authors in [35] and [36] have expressed skepticism.

We examined animal communication of fear and safety related to foraging behavior and fitness using EcoSim. The individuals were given the ability to communicate their feelings of fear (potentially due to nearby predators) to their neighbors and then we observed the effects of the same on the population of prey, to shed some light on the ongoing debates. We also utilized a wrapper method combining c4.5 algorithm with VAE-EDA-Q, as a combinatorial optimizer for feature selection, to determine rules that specified the conditions related to alarm communication.

Bibliography

[1] Anil K Seth. The ecology of action selection: Insights from artificial life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1545–1558, 2007.

[2] John R Krebs and Nicholas B Davies. *Behavioural ecology: an evolutionary approach*. John Wiley & Sons, 2009.

[3] Morteza Mahayekhi, Abbas Golestani, Yasaman Farahani, and Robin Gras. n Enhanced Artificial Ecosystem: Investigating Emergence of Ecological Niches. In *Artificial Life Conference Proceedings 14*, pages 693–700. MIT Press, 2014.

[4] Morteza Mashayekhi. Individual-Based Modeling and Data Analysis of Ecological Systems Using Machine Learning Techniques. 2015.

[5] Carlo Ricotta. From theoretical ecology to statistical physics and back: self-similar landscape metrics as a synthesis of ecological diversity and geometrical complexity. *Ecological Modelling*, 125(2-3):245–253, 2000.

[6] Didier Devaurs and Robin Gras. Species abundance patterns in an ecosystem simulation studied through Fisher's logseries. *Simulation Modelling Practice and Theory*, 18(1):100–123, 2010.

[7] Geoffrey Keppel and Sheldon Zedeck. *Data analysis for research designs*. Macmillan, 1989.

[8] Cathy O'Neil and Rachel Schutt. *Doing data science: Straight talk from the frontline*. " O'Reilly Media, Inc.", 2013.

[9] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[10] Mark A Hall and Lloyd A Smith. Practical feature subset selection for machine learning. 1998.

[11] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[12] Shumeet Baluja. Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, 1994.

[13] Heinz Mühlenbein and Gerhard Paass. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*, pages 178–187. Springer, 1996.

[14] Pedro Larrañaga and Jose A Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.

[15] Iñaki Inza, Pedro Larrañaga, Ramón Etxeberria, and Basilio Sierra. Feature subset selection by Bayesian network-based optimization. *Artificial intelligence*, 123(1-2):157–184, 2000.

[16] Marwa Khater and Robin Gras. Adaptation and genomic evolution in ecosim. In *International Conference on Simulation of Adaptive Behavior*, pages 219–229. Springer, 2012.

[17] Elham Salehi and Robin Gras. Efficient EDA for large opimization problems via constraining the search space of models. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 73–74. ACM, 2011.

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[19] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic back-propagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[20] Robin Gras, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life*, 15(4):423–463, 2009.

[21] Johan Dannewitz, Gregory E Maes, Leif Johansson, Håkan Wickström, Filip AM Volckaert, and Torbjörn Järvi. Panmixia in the European eel: a matter of time.... *Proceedings of the Royal Society B: Biological Sciences*, 272(1568):1129–1137, 2005.

[22] David J Hosken and Clarissa M House. Sexual selection. *Current Biology*, 21(2):R62–R65, 2011.

[23] Anders Pape MÖller and Rauno V Alatalo. Good-genes effects in sexual selection. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1414):85–91, 1999.

[24] John A Byers and Lisette Waits. Good genes sexual selection in nature. *Proceedings of the National Academy of Sciences*, 103(44):16343–16345, 2006.

[25] Sourodeep Bhattacharjee, Brian MacPherson, and Robin Gras. A comparison of sexual selection versus random selection with respect to extinction and speciation rates using individual based modeling and machine learning. *Ecological Complexity*, 36:126–137, 2018.

[26] Karl Von Frisch. The dance language and orientation of bees. 1967.

[27] Delbert L Smee and Marc J Weissburg. Hard clams (Mercenaria mercenaria) evaluate predation risk using chemical signals from predators and injured conspecifics. *Journal of chemical ecology*, 32(3):605–619, 2006.

[28] Matthew D Mitchell, Peter F Cowman, and Mark I McCormick. Chemical alarm cues are conserved within the coral reef fish family Pomacentridae. *Plos one*, 7(10):e47428, 2012.

[29] R Inta, TA Evans, and JCS Lai. Effect of vibratory soldier alarm signals on the foraging behavior of subterranean termites (Isoptera: Rhinotermitidae). *Journal of economic entomology*, 102(1):121–126, 2009.

[30] A Randall Hughes, David A Mann, and David L Kimbro. Predatory fish sounds can alter crab foraging behaviour and influence bivalve abundance. *Proceedings of the Royal Society B: Biological Sciences*, 281(1788):20140715, 2014.

[31] Steven L Lima and Peter A Bednekoff. Temporal variation in danger drives antipredator behavior: the predation risk allocation hypothesis. *The American Naturalist*, 153(6):649–659, 1999.

[32] Andrew Sih and Thomas M McCarthy. Prey responses to pulses of risk and safety: testing the risk allocation hypothesis. *Animal Behaviour*, 63(3):437–443, 2002.

[33] Reehan S Mirza, Alicia Mathis, and Douglas P Chivers. Does temporal variation in predation risk influence the intensity of antipredator responses? A test of the risk allocation hypothesis. *Ethology*, 112(1):44–51, 2006.

[34] Stefanie Slos and Robby Stoks. Behavioural correlations may cause partial support for the risk allocation hypothesis in damselfly larvae. *Ethology*, 112(2):143–151, 2006.

[35] Maud CO Ferrari, Andrew Sih, and Douglas P Chivers. The paradox of risk allocation: a review and prospectus. *Animal Behaviour*, 78(3):579–585, 2009.

[36] Guy Beauchamp and Graeme D Ruxton. A reassessment of the predation risk allocation hypothesis: a comment on Lima and Bednekoff. *The American Naturalist*, 177(1):143–146, 2010.

Chapter 2

Background

2.1 Individual Based Modeling

2.1.1 Need for Individual Based Modeling

Models, in general, are simplified representations of real systems and universally share the challenge of proving their predictive capabilities [1]. Ecological models, however, have the unique requirement of depicting relevant spatial and temporal scales in tandem with a multitude of processes representative of the system observed [2]. Ecological models support environmental decision making in ways that corresponding experiments cannot as conclusions drawn from descriptive studies have the potential of failing to fully represent processes [3]. While the study of ecology involves entire populations, communities, and ecosystems, the properties of the system are principally determined by the properties and behavior of the individuals from which they are composed. Therefore, individuals are the foundation of ecological models [4].

Authors in [4] extrapolate this further by drawing comparison to the field of physics: many physics topics may be addressed without referring distinctly to atoms, despite the concept that atomic properties and interactions determine the foundations

of matter. This illuminates a key difference – individual organisms have biotic properties that atoms lack. The principles of life cycles, growth, development, reproduction, and death are conserved throughout ecological systems despite the transient nature of the individuals to which they belong. Additionally, individuals modify their environments through interaction with resources; even within the same species, differences allow individuals to modify the environment in distinctive ways. Most important, though, is the concept of adaptation wherein individuals are able to grow, mature, obtain resources, reproduce, and interact depending upon intrinsic and extrinsic factors [4]. This notion highlights the difference between atomic theory and ecology – individual organisms are adaptive because their response to biotic and abiotic factors determines their ability to pass their genes on to future generations (fitness). Furthermore, fitness-seeking adaptation does not occur to advance the population as a whole; behavioral adaptation occurs at the level of the individual.

2.1.2 What is Individual Based Modeling?

Individual based models are capable of handling the high degree of complexity in the representation of individuals and interactions among individuals. This approach (also known as agent based models) is described to simulate populations and systems with respect to each individual organism [5]. Individuals have their own set of state variables, which include spatial location and physiological or behavioral traits. Attributes such as growth, habitat selection, foraging, reproduction, and dispersal are able to differ among individuals and change over time [6].

Traditional differential equation population models use top-down population parameters like birth rates and death rates. Also known as EBM (Equation-based modeling), traditional methods use equation-based simulations and are unable to accurately depict life histories of individuals [7]. Individual based models instead employ

a bottom-up approach where interactions among discrete, autonomous individuals, as well as among their abiotic environments, drive the population level behaviors [8] and [9]. These interactions are the foundation of emergent properties like species distribution at the population and ecosystem levels [7]. By employing discrete units, the incorporation of individual level mechanisms can be represented. This is in direct contrast to traditional models where complexity and interactions cannot be represented to this degree. IBM makes possible the examination of variation of individuals at life cycle stages, variation among individuals, local interactions among individuals, and adaptive behaviors such as energy budgets and physiology [6]. Authors in [7] suggest that the distinguishing characteristics amount to four keys: degree of complexity of individual life cycles, variation of resources used, quantities measured in discrete numbers versus real numbers, and variation among individuals of the same age.

IBMs account for differences among individuals by categorizing increasing levels of mechanistic details along five axes; this embodies the spirit of inclusion of variation among individuals, while also reflecting a balance between parsimony and detail. Categories of axes include: spatial variability, life cycle and ontogenetic development, phenotypic variability, cognitive variability, and genetic variability/evolution [8]. Variability in space refers to the physical space in which interactions occur. Within IBM, predator-prey models may have metapopulation or reaction-diffusion foundations. Space variability under metapopulation models is reflected by taking into consideration the distances between patches, while population density spread through space is represented by reaction-diffusion models. Life cycle and ontogenetic development can be differentially represented in terms of demographics, spatial structure, and temporal variability. Within matrix models, details such as these would be organized into a large number of subclasses; therefore, the use of IBMs is highly expedient. Phenotypic variability is represented with a greater magnitude through

IBMs as the unique experiences of each individual, in addition to variation caused by ontogenetic changes, can be used to determine trajectory. Using plants as an example, it is possible to illustrate how the inclusion of details such as soil, water, nutrients, and light gives rise to an incredibly more precise history. Cognitive variability shares a similar foundation with phenotypic variability within individual based modeling – experience and learning are derived from individual experiences. Memories of past experiences are considered to be an internal state and give rise to learning, which can arise from the environment of another organism. Genetic variability and evolution studies have classically focused on the individual. Mutations, genetic drift, and founder effects are examples of evolutionary genetic concepts that involve a small number of individuals yet create a profound effect. IBM also uses an individualistic foundation, which makes the concept of stochasticity significant. IBMs have a higher degree of flexibility than classical models, which allows for the analogous representation of true population change.

2.1.3 Harnessing the power of IBM for Ecological and Evolutionary Processes

Within individual-based models, two models exist to serve two separate aims [10]. Pragmatic models center on the notions that representing individuals explicitly is essential in the simulation of specific populations, communities, or ecosystems and that simple mathematical models cannot address these problems. Queries addressed within these models are generally management related. Paradigmatic models, on the other hand, address queries regarding the underlying cause of ecological phenomena. These models assume a paradigm shift from traditional physics style modeling to models based on emergence and evolutionary principles is necessary [6], [11], and [12]. Six major ecological processes have been highlighted as main categories served by

paradigmatic IBMs [8].

i Local interactions and movement - The examination of movement through space entails modeling a vast grouping of detailed active movement behaviors from local interaction with animals and landscapes to development of home ranges.

ii Formation of patterns among individuals - Formation of patterns among individuals describes the study of how social forces, environmental factors, and individual decisions give rise to swarms or other aggregations.

iii Interactions of exploitative species - When considering exploitative species, spatial movement patterns have been shown to radically affect the stability of the interaction through the diffusion and mixing of populations in predator-prey and host-parasitoid interactions.

iv Community dynamics and local competition - When examining sessile organisms, focus on emergent phenomena and community dynamics are of great interest and are studied using grid cells or continuum models. Competition-colonization trade off, effects of conspecific density, and niche differences are all able to interact in spatial context to illustrate factors that control species richness and diversity.

v Population dynamics with focus on foraging and bioenergetics - Individual variation as related to population dynamics is studied by examining foraging and bioenergetics. Frequency-dependent selection and variance in growth are two parameters of foraging models that may be used to highlight a predator's preference (which would examine the effects in variation of prey). Foraging decisions in birds have been examined with complex models that use artificial life approaches. These methods have shown that spatial memory is a key to successful foraging as it relates to resource distribution. Foraging models in combination with physiological facilities/constraints are used to study bioenergetics as they consider time and energy budgets to make precise predictions in regards to individual decisions.

vi Evolutionary process - Evolutionary process modeling seeks to answer a multitude of queries, which range from comparisons of trait values to genetic algorithms examining optimal traits for foraging, avoidance of predators, reproduction, and dispersal. While they do not replicate mechanisms of evolution, these models do conserve genetic diversity and have even predicted settings in which polymorphisms could be maintained within populations.

Through a radical departure from classical mathematical approaches to ecological theory, IBM has established a new philosophical paradigm. By requiring the inclusion of individual detail into models, the term IBM is considered tantamount to an explicit examination of individuals and their complex responses to their environments [8]. The rule-based simulations employed by IBM are optimal for responses such as phenotypical change and learning than are their mathematical model counterparts. Population and community level behaviors arise from adaptive behaviors of individuals. Future developments will focus on further sophistication of representations of internal states and increased autonomy. This will shed light on the decisions of individuals and on behaviors such as mating. Pattern oriented modeling will allow the comparison of model behavior to natural systems [13]. Authors in [9] uphold that a formally documented model is sufficient in terms of rigor and that mathematical notation is not a requisite. The future development of individual based modeling will play a great role in paradigmatic ecology and may provide understanding for the basis of evolution [10].

Bibliography

[1] Edward J Rykiel Jr. Modeling agroecosystems: Lessons from ecology. *Agricultural Ecosystems: Unifying Concepts, R. Lowrance, B. Stinner, and G. House, eds. New York: Wiley*, pages 157–178, 1984.

[2] Amelie Schmolke, Pernille Thorbek, Donald L DeAngelis, and Volker Grimm. Ecological models supporting environmental decision making: a strategy for the future. *Trends in ecology & evolution*, 25(8):479–486, 2010.

[3] Jacqueline Augusiak, Paul J Van den Brink, and Volker Grimm. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecological Modelling*, 280:117–128, 2014.

[4] Volker Grimm and Steven F Railsback. *Individual-based modeling and ecology*, volume 8. Princeton university press, 2013.

[5] H Van Dyke Parunak, Robert Savit, and Rick L Riolo. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 10–25. Springer, 1998.

[6] Donald L DeAngelis and Volker Grimm. Individual-based models in ecology after four decades. *F1000prime reports*, 6, 2014.

[7] Brian MacPherson and Robin Gras. Individual-based ecological models: Adjunctive tools or experimental systems? *Ecological Modelling*, 323:106–114, 2016.

[8] Donald L DeAngelis and Wolf M Mooij. Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol. Evol. Syst.*, 36:147–168, 2005.

[9] V Grimm and SF Railsback. Individual-based modeling and ecology, 480 pp, 2005.

[10] Volker Grimm. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological modelling*, 115(2-3):129–148, 1999.

[11] Adam Lomnicki. *Population ecology of individuals*. Princeton University Press, 1988.

[12] Janusz Uchmański. Differentiation and frequency distributions of body weights in plants and animals. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 310(1142):1–75, 1985.

[13] Volker Grimm, Karin Frank, Florian Jeltsch, Roland Brandl, Janusz Uchmański, and Christian Wissel. Pattern-oriented modelling in population ecology. *Science of the Total Environment*, 183(1-2):151–166, 1996.

2.2 Estimation of Distribution Algorithms: Machine Learning Approaches

2.2.1 Introduction to Machine Learning Based EDA

Estimation of Distribution Algorithms (EDA) [1], [2] are algorithms that employ meta-heuristics to aid in combinatorial and continuous non-linear optimization [3]. Within EDA, an efficient probabilistic model guides the search towards solutions of higher fitness, compared to solutions of the previous generation. The model is generated by encoding the probability distribution of admissible solutions based on their fitness values [4]. EDA performs a random sampling of the probabilistic model to yield new solutions representative of the population as a whole, while maintaining a similar or desirably better solution quality than the present solutions of the population. In order to discover the optimal solution, EDA initiates a population of uniformly and randomly generated candidate solutions. This population of candidate solutions is iteratively improved upon over consecutive generations. EDA estimates the possibility of a candidate to lead to an optimal offspring over subsequent generations while also discerning the hidden interdependencies between the constituent features in the candidate.

Models based on factorizations of univariate frequencies, such as the Univariate Marginal Distribution Algorithm (UMDA) [1], produce estimates from a given population very quickly but fail to model the interdependencies between decision variables. Hence, EDAs that model univariate frequencies independently cannot solve complex, real-world problems. There are alternative approaches that consider interdependencies between variables. Bayesian Networks, as in the Bayesian Optimization Algorithm (BOA) [5] [6], are complex models that learn Bayesian networks as a probabilistic

model of the population. These models solve problems composed of subproblems of overlapping variables that cannot be decomposed into independent subproblems due to inter-dependency. This method facilitates the solving of very complex problems; however, estimating the model from a population can be very computationally demanding. The production of a probabilistic model that is both flexible and efficient in terms of estimation and sampling is of prime importance. Parallel research in the field of Machine Learning algorithms shares the same focus, for example unsupervised machine learning algorithms for generative neural networks [3]. These algorithms are able to accept high dimensional data as input, from which they are able to learn complex patterns. Generative neural networks produce models from a population of samples in an unsupervised manner and are capable of generating from them completely new solutions having the same likelihood distribution. This makes them very useful for EDA applications as they perform better exploration of the search space while using less time and computational resources. These generative neural network models can be stacked upon each other in layers and provide building blocks for use in deep learning.

2.2.2 Restricted Boltzmann Machine EDA

A Restricted Boltzmann machines based method (RBM) [7] has been proposed by authors in [3] to model the population of EDAs. RBMs are stochastical neural networks with a basic structure consisting of a visible or input layer of nodes (V) and a hidden or output layer (H) called feature detectors; the hidden layer models patterns in the data. During the sampling step, a particular configuration of the network is sampled with the joint probability distribution P(V, H). The authors compared RBM-EDA to BOA on OneMax, concatenated (4, 5 and 6) traps, and NK Landscapes for various problem sizes. Their results shown that a greater number of fitness evalua-

tion is required for RBM-EDA than BOA, which would indicate that the population model of RBM-EDA was less accurate than the statistical model of BOA. However, the authors also stipulated that the quality of the model was compensated for by the shorter model building time for RBM-EDA. They also found that the CPU time for RBM-EDA grew at a slower rate than BOA with an increase in problem sizes; and for difficult problems, the performance of RBM-EDA was similar or superior to BOA.

2.2.3 Denoising Autoencoder EDA

Autoencoders [8–10] are neural networks that learn an abstract representation of the input through an encoder model and later performing a reconstruction step using a decoder. This process converts the representation to an output that approximates the distribution of training examples. Autoencoders have been utilized as generative modeling techniques in dimensionality reduction and for feature learning applications [11].

To be conducive to dimensionality reduction, it is imperative that autoencoders learn to avoid generating the same output as the input provided; this is accomplished by using regularization to restrict over-fitting on the training data [12]. One method used to prevent such over-fitting is Denoising Autoencoders (DAE) [13]. This method intentionally corrupts the input or training sample with a certain type of noise (uniform, Gaussian, etc.) and trains the denoising autoencoder to filter out the noise in a denoising layer. Thereafter, the output is reconstructed from a “cleaned" input, which makes denoising autoencoders robust to noise in the training samples. Denoising Autoencoder EDA (DAE-EDA) [14] uses denoising autoencoders as generative models for EDA. DAE-EDA was tested on concatenated deceptive traps, NK Landscapes, and HIFF functions, and its performance was compared to BOA and RBM-EDA. DAE-EDA was found to be much faster than BOA and RBM-EDA, but at the cost of

a reduced optimization score. In some cases, DAE-EDA was able to reach quicker convergence in term of number of fitness evaluations, but in most other cases it required larger population sizes to achieve guaranteed convergence.

2.2.4 Deep Boltzmann Machines EDA

Deep Boltzmann Machines (DBM) [15] are deep neural networks consisting of multiple layers of hidden neurons, and can thus be used to capture the model of the population at increasing layers of abstraction. Each hidden layer provides one additional layer of abstraction capable of representing patterns or features in the data. Authors in [16] used DBM with EDA based on the promise shown by the results in RBM-EDA and on the success of deep learning models. Following similar experimental benchmarks as in RBM, the authors found that while DBM-EDA was computationally less expensive than BOA, the quality of the solutions was inferior to BOA in cases of multi modal problems beyond trap-5. Furthermore, the authors specified that given the efforts needed to train the multi-layer DBM-EDA, it was not feasible to be used in a noisy training set. DBM-EDA was not able to find the global optimum in complex problems when allotted the same population size as BOA.

2.2.5 Generative Adversarial Network EDA

An alternative generative approach for EDA modeling was suggested by the author in [17], where Generative adversarial networks [**?**] were used to model the population [17] (GANEDA). GANs are generative neural networks that estimate the distribution of data using two components – the generator and the discriminator. They are trained separately and iteratively. The generator produces solutions that represent the distribution of the population, while the discriminator tries to distinguish the solutions produced by the generator from the rest of the population. The task

of the generator is to produce new solutions that very closely reflect the distribution of the population such that the discriminator cannot detect the new solution as a synthesized outlier. Authors in [17] tested GAN-EDA on one-max problems, concatenated trap functions, and NK Landscapes. GAN EDA displayed lower performance than both DAE-EDA and BOA for the test problems presented and in most cases was unable to find the global optimum.

2.2.6 Variational Autoencoder EDA

Variational Autoencoders (VAE) [18] [19] improve upon traditional autoencoders in their explorative capabilities by generating data from continuous latent spaces representation, which allows for smooth interpolation and randomized sampling by expanding the latent search space horizon beyond what the training samples represent. While traditional autoencoders are able to introduce minor random variations in the input provided, VAEs allow a greater degree of control of the direction and magnitude of the variation produced in the output by the hidden model [20]. Autoencoders described in previous sections differ from VAE in that standard autoencoders do not have a probabilistic basis, whereas VAEs allow stochastic generative modeling with tunable probability parameters [21]. Authors in [22] used VAE as a generative modeling technique for EDA population (VAE- EDA). The presence of continuous latent spaces within VAE allows for easy interpolation and random sampling. A first version of VAE-EDA was proposed by authors in [22]; a randomized population was generated initially, from which the fittest subset of individuals were identified using Tournament selection. From the fittest individuals, a Gaussian probabilistic VAE model was learned, which was then sampled to derive offspring candidates for the next generation. The offspring were combined with parents, and the selection procedure was carried out again until the global optimum was found.

The authors proposed two extensions of VAE-EDA. The first extension, named Extended VAE (E-VAE), contained a second decoder along with the standard encoder-decoder. The second decoder was used to learn to predict the fitness value of the output. The second extension proposed by the authors in [22] is titled Conditioned, Extended VAE (CE-VAE) and proposed to use the predictor to explicitly sample the solutions with best-predicted fitness. In this extension, the predictor accepted the output of the encoder (same as the decoder that outputs the off-spring), and the predictor and decoder were trained simultaneously. This predictor demonstrated efficacy as a regularizer component for the latent representation and also demonstrated potential to be used as a surrogate fitness function in situations where the actual fitness computation required a great amount of time. Substituting the surrogate fitness function was found to improve the overall performance of the algorithm in cases such as this. This method, however, was not tested on any benchmark problems or compared to any state-of-the-art algorithms. Authors in [22] merely compared the relative performance of three methods on a simplified protein folding problem that had a single objective.

Bibliography

[1] Heinz Mühlenbein and Gerhard Paass. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*, pages 178–187. Springer, 1996.

[2] Pedro Larrañaga and Jose A Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.

[3] Malte Probst, Franz Rothlauf, and Jörn Grahl. Scalability of using Restricted Boltzmann Machines for combinatorial optimization. *European Journal of Operational Research*, 256(2):368–383, 2017.

[4] Martin Pelikan, Mark W Hauschild, and Fernando G Lobo. Estimation of distribution algorithms. In *Springer Handbook of Computational Intelligence*, pages 899–928. Springer, 2015.

[5] Martin Pelikan, David E Goldberg, and Erick Cantu-Paz. Linkage problem, distribution estimation, and Bayesian networks. *Evolutionary computation*, 8(3):311–340, 2000.

[6] Martin Pelikan, David E Goldberg, and Erick Cantú-Paz. BOA: The Bayesian optimization algorithm. In *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*, pages 525–532. Morgan Kaufmann Publishers Inc., 1999.

[7] Paul Smolensky. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Information Processing in Dynamical Systems: Foundations of Harmony Theory. *MIT Press, Cambridge, MA, USA*, 15:18, 1986.

[8] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[9] Y-lan Boureau, Yann L Cun, et al. Sparse feature learning for deep belief networks. In *Advances in neural information processing systems*, pages 1185–1192, 2008.

[10] Yoshua Bengio et al. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[11] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[12] Yoshua Bengio, Li Yao, Guillaume Alain, and Pascal Vincent. Generalized denoising auto-encoders as generative models. In *Advances in Neural Information Processing Systems*, pages 899–907, 2013.

[13] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.

[14] Malte Probst. Denoising autoencoders for fast combinatorial black box optimization. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, pages 1459–1460. ACM, 2015.

[15] Ruslan Salakhutdinov and Geoffrey Hinton. Deep Boltzmann Machines. In David van Dyk and Max Welling, editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 448–455, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 2009. PMLR.

[16] Malte Probst and Franz Rothlauf. Deep boltzmann machines in estimation of distribution algorithms for combinatorial optimization. *arXiv preprint arXiv:1509.06535*, 2015.

[17] Malte Probst. Generative adversarial networks in estimation of distribution algorithms for combinatorial optimization. *arXiv preprint arXiv:1509.09235*, 2015.

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[19] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic back-propagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[20] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

[21] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2:1–18, 2015.

[22] Unai Garciarena, Roberto Santana, and Alexander Mendiburu. Expanding variational autoencoders for learning and exploiting latent representations in search distributions. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 849–856. ACM, 2018.

2.3 Wrapper Based Feature Selection and EDA Approaches

When machine learning algorithms are applied on high dimensional data, an issue known as curse of dimensionality arises, where as the number of dimension increases, the volume of the data space increases and the amount of data available becomes sparser [1]. Sparse data adverse affects the efficacy and the solution quality of several machine learning algorithms [2] for example for classification models that work on grouping instances based on their similarities to come up with a model, struggle to make such groupings when due to extraneous features they mind most instances to be dissimilar. Sometimes the presence of a large number of features causes learning models to overfit the training data which degrades the solutions given by the model when tested on novel data. Not only that, working with high dimensional data increases memory requirements and puts a higher demand on computational costs for machine learning algorithms [2].

A class of machine learning algorithms exist to tackle the issues arising from high dimensional data. These algorithms are known as Dimensionality reduction algorithms and are categorized into two groups: Feature Selection and Feature Extraction. In Feature Extraction, subsets of higher dimensional data are combined linearly or non-linearly and the output is mapped to a new feature space. The features in the new feature space thus created have lower number of dimensions than the original feature space. On the other hand, feature selection directly choses a subset of features from the original feature space to lower the dimensionality [3] [4].

Both feature selection and feature extraction are considered appropriate tools to reduce features in high dimensional data leading to better generalization of models, while also decreasing computational costs and memory requirements. Feature ex-

traction produces high level features by combining features from the original feature space. These high-level features are making it difficult to interpret the resulting model built from these high-level features. Hence, in many cases feature selection becomes a preferred choice as it retains the relevant original features and only removes the unnecessary features, thereby maintaining the ease of interpretation of the models generated from the subset [2]. Even in cases where the original number of features are not too high, feature selection is often employed to reduce computational costs and to improve model quality. Moreover, real-world data often contains a lot of noisy features that are irrelevant and sometimes redundant. Removal of such features improve the performance of machine learning algorithms while improving computational efficiency. In the context of this body of research, we will focus our discussion on feature selection methods.

Features selection techniques are classified into three categories based on the level of coupling with other machine learning algorithms (such as classification algorithms). Thus, depending on the method the feature selection search is performed along with the construction of classification model, the feature selection algorithms can be classified as: Filter methods, Wrapper methods and Embedded Methods [5].

Filter methods analyze only the intrinsic properties of the data and usually work by assigning a feature relevance score to the individual features in the feature space. Based on a certain threshold, the lower scoring features are removed from the feature space. The remaining higher scoring features are fed into classification algorithms for further processing. Hence filter methods are computationally inexpensive in most cases and work independently of the classification algorithms that consume the data.

In contrast to filter methods, Wrapper methods integrate the feature search step within the model hypothesis search of the classification method. In wrapper methods, a search method is defined that selects a subset of features in the feature space. A spe-

cific subset of features produced is evaluated by using the feature subset to train and test a specific classification algorithm and model. Hence this method is very tightly coupled with the specific problem domain and classification method employed. A major advantage of wrapper method is that they take into account the interaction between model search and feature subset search and can also discover feature dependencies. Disadvantages of wrapper method is that they tend to overfit the data and are computationally expensive.

Estimation of Distribution Algorithms (EDA) [6], [7] have been used for feature subset search in wrapper methods in previous research. In [8] the authors proposed a novel algorithm - Feature Subset Selection by Estimation of Bayesian Network Algorithm (FSS - EBNA) which uses Estimation of Bayesian Network Algorithm [9] which follows EDA paradigm for feature subset selection. EDAs evolves the population by altering probability distribution of the highest fitness candidates in the population of candidates in each iteration of the search. In EBNA, the evolution of the model is performed by a Bayesian network working in tandem with a local search method. Authors claimed FSS-EBNA to be a computationally efficient wrapper method that can be successfully applied to any problem where specific domain knowledge is not available and where number of samples available is very low.

In [10] the authors used wrapper methods with EDAs to classify cancerous genes in gene expression datasets. They successfully used naïve Bayes classifier with EDA as a wrapper, to considerably reduce the number genes in their classification model, leading to a concise model that is easy to interpret which was a critical requirement for their problem. EDAs, specifically Population based Incremental Learning [11] along COMIT which is a dependency tree based EDA, were used as wrappers in [12] for predicting survival rates of cirrhotic patients.

EDA methods were also applied a wrapper technique in [13] on results obtained

from EcoSim, an artificial ecosystem simulator [14]. One of the objectives of this study was to discover the genes that have a stronger influence on fitness of individuals. In order to achieve this, the authors wrapped CMSS-BOA [15] for feature search on to Random Forests [16] for classification. CMSS-BOA does not restrict a fixed upper bound on the number of variables on which another variable can have some dependency with, leading to discovery of intricate and highly relevant interdependencies between variables. Each subset of interdependent variables is encoded as a string of bits. The subset that maximizes the Area under ROC curve (AUC) obtained by Bayesian network classifier is selected.

In conclusion, feature selection is an effective and efficient tool to address the problems associated with high dimensional datasets leading to concise and interpretable machine learning models which can be built in reasonable amount of computational time and resources. Wrapper methods integrate the feature subset search with classifier model search leading to selection of subset of features that are guaranteed to result in better classification models. Estimation of Distribution algorithms being computationally efficient tools for continuous and non-linear combinatorial optimizations have proven to be a good choice for use as a wrapper method in feature subset search and selection in previous research.

Bibliography

[1] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

[2] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6):94, 2018.

[3] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti A Zadeh. *Feature extraction: foundations and applications*, volume 207. Springer, 2008.

[4] Huan Liu and Hiroshi Motoda. *Computational methods of feature selection*. CRC Press, 2007.

[5] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[6] Heinz Mühlenbein and Gerhard Paass. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*, pages 178–187. Springer, 1996.

[7] Pedro Larrañaga and Jose A Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.

[8] Iñaki Inza, Pedro Larrañaga, Ramón Etxeberria, and Basilio Sierra. Feature subset selection by Bayesian network-based optimization. *Artificial intelligence*, 123(1-2):157–184, 2000.

[9] Ramon Etxeberria. Global optimization using Bayesian networks. In *Proc. 2nd Symposium on Artificial Intelligence (CIMAF-99)*, 1999.

[10] Rosa Blanco, Pedro Larrañaga, Iñaki Inza, and Basilio Sierra. Gene selection for cancer classification using wrapper approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(08):1373–1390, 2004.

[11] Shumeet Baluja. Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, 1994.

[12] Iñaki Inza, Marisa Merino, Pedro Larrañaga, Jorge Quiroga, Basilio Sierra, and Marcos Girala. Feature subset selection by genetic algorithms and estimation of distribution algorithms: A case study in the survival of cirrhotic patients treated with TIPS. *Artificial Intelligence in Medicine*, 23(2):187–205, 2001.

[13] Marwa Khater and Robin Gras. Adaptation and genomic evolution in ecosim. In *International Conference on Simulation of Adaptive Behavior*, pages 219–229. Springer, 2012.

[14] Robin Gras, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life*, 15(4):423–463, 2009.

[15] Elham Salehi and Robin Gras. Efficient EDA for large opimization problems via constraining the search space of models. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 73–74. ACM, 2011.

[16] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Chapter 3

Estimation of Distribution using Population Queue based Variational Autoencoders

3.1 Introduction

Estimation of Distribution Algorithms (EDA) [1–4] are population-based stochastic combinatorial optimization algorithms that attempt to explore a solution space to find an optimum solution by creating probabilistic models of promising solutions and creating new samples from the probabilistic model. The algorithms typically try to generalize from the current solution space sample (current population) to build a probabilistic model; this model is then sampled to create the next generation of solutions that are evaluated with respect to the function to optimize (the fitness function). The best new solutions are selected and are recombined with the parent population (see figure 5.1 adapted from [5]). The population model is improved at every iteration of EDA to yield the next generation of solutions that could be

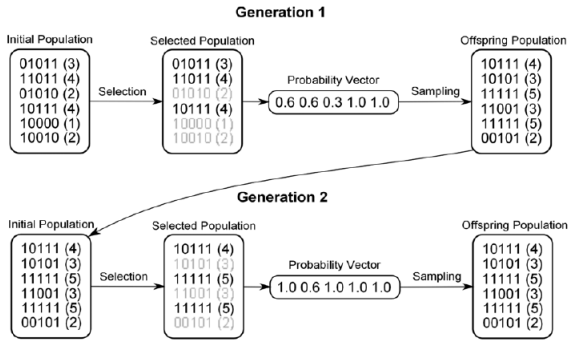


Figure 3.1: EDA Procedure: The initial population is assigned a fitness value by application of a fitness function on the candidate, and a selection step is applied to pick the fittest candidates out of the parent pool. A simple probabilistic model is created from the fittest parents which is then sampled to produce next generation of candidates. [5]

potentially closer to the global optima.

In recent studies, Machine Learning techniques have been employed to build population models with considerable success. Restricted Boltzmann Machine (RBM) [6] comprised of a stochastic neural network was used as population model [7] and displayed a considerably less computational time requirement than BOA. In another work, the authors implemented Deep Boltzmann Machine (DBM) to generate the next batches of solutions [8]. Autoencoders [9, 10] have also been used to model the population of promising solutions [11].

In this paper we propose a novel extension for EDA that uses Variational Autoencoders [12] - Variational Autoencoder EDA (VAE-EDA) [13]. Variational autoencoders are expected to yield models that explore the solution space more optimally as they model a probabilistic distribution of the latent space (its average and standard deviation on the assumption that the latent variables follow a Gaussian distribution) that would better capture the properties of the objective function, which allows for the smoothing of the data generation process [13]. The extension of VAE-EDA (VAE-EDA-Q) proposed in this paper smooths the update of the VAE-EDA model to avoid

premature convergence to a sub-optimal region of the solution-space by learning the model on a set of candidate solutions generated on a range of previous generations (modeled as a queue of populations) of the VAE-EDA process. Another extension proposed in this work uses Adaptive Variance Scaling (AVS) [14] to dynamically control the rate of exploration of the latent space by using a coefficient multiplied to the variance of sampling.

The success of machine learning heavily relies on human machine learning experts who are responsible for feature selection, workflow design, selection and design of machine learning models, and hyper-parameters. There is a growing demand for self-contained machine learning methods that can be used within a variety of domains without the necessity of the involvement of machine learning human experts. To this effect, [15] a new field of research targeting progressive automation of machine learning, known as AutoML, is being pursued. The term AutoML encompasses all aspects related to automating the process of machine learning beyond model search, hyper-parameter optimization, and algorithm selection and includes representation learning and automatic feature extraction, automatically applying algorithms to a given problem, automatic detection of skewed data and missing values, etc. [15].

Convolutional Neural Networks [16] [17] (CNNs) are neural networks that aid in processing multi-dimensional array representations of complex data such as a color image stored as three two dimentional arrays, for example [18]. Various techniques have been previously employed to automatically generate CNNs. These methods can be broadly categorized as evolutionary algorithms [19] or reinforcement learning [20] based methods. Evolutionary algorithm approaches perform by altering a population of promising solutions using evolutionary strategies along with specific heuristics to quickly find the best candidate. Prominent contributions include Genetic CNN [21], CNN-GA [22] that utilizes genetic algorithms to discover CNN architectures, Large-

scale Evolution [23], Hierarchical Evolution [24] that utilizes hierarchical representation, and Catesian Genetic Programming method (CGP) [25] CNN. Reinforcement learning approaches employ a form of reward-penalty strategy where a reward is given for finding solutions close to the optimum and a penalty is imposed when the solution is farther from optimum. Other algorithms that fall into this category include neural architecture search (NAS) method [26], MetaQNN that employs meta modeling [27], the efficient architecture search (EAS) method [28], and block design method (Block-QNN-S) [29].

Empirically, the algorithms discussed above have demonstrated promising results on image recognition and classification problems when tested on CIFAR10 and CIFAR100 [30], which are considered as benchmark datasets for image classification problems. However, due to the challenges involved in the automatic generation of CNNs (including the heavy demand on computational resources and computational time) some form of manual intervention is still necessary for most algorithms within both categories. Additionally, the resulting classification accuracy does not perform as well as the state-of-the-art. Fully automatic algorithms mentioned previously include CNN-GA, Large Scale Evolution, Meta CNN, NAS, and CGP CNN. Semi-automatic algorithms include Genetic CNN, Block-QNN-S, EAS, and Hierarchical Evolution.

In this paper, we have used VAE-EDA-Q with AVS as an AutoML algorithm to discover architectures for CNNs and tested the CNNs generated as compared to the results of other state of the art algorithms.

3.2 Background and Related Work

3.2.1 Bayesian optimization algorithm and improvements

Bayesian optimization algorithm (BOA) [31] [32] is a state-of-the-art EDA capable of tackling many difficult optimization problems. It uses Bayesian networks to model the population and, hence, solves many problems that are difficult to decompose into separable subproblems. The possible configuration of a Bayesian Network grows exponentially with the number of decision variables. Hence, a greedy heuristic is used to build a network from an empty network consisting of nodes only. Some real world problems are hierarchically decomposable, and hBOA [33] algorithm tackles such problems by creating compact Bayesian networks with local structures to allow complex networks to be learned [5]. Additionally, hBOA uses restricted tournament replacement, which is a niching technique that attempts to match similar solutions against each other rather than competing dissimilar solutions.

Constrained Model Search Space BOA (CMSS-BOA) [34], on the other hand, successfully improves on the computational efficiency of BOA by constraining the search space using a heuristic called max-min parent children (MMPC) [35] and then performing hill-climbing on the search space. CMSS-BOA was compared to BOA on benchmark problems such as OneMax and concatenated k-trap function. CMSS-BOA was found, in most cases, to be able to converge earlier than BOA. However, the average fitness of BOA population was slightly higher than CMSS-BOA.

3.2.2 Deep Boltzmann Machine based EDA

Deep Boltzmann Machines (DBM) [36] are deep neural networks consisting of multiple layers of hidden neurons, and, hence, can be used to capture the model of the population at increasing layers of abstraction. Each hidden layer provides one addi-

tional layer of abstraction capable of representing patterns or features in the data. The authors in [8] used DBM with EDA based on the promise shown by the results in RBM-EDA and based on the success of deep learning models. The authors found that while DBM-EDA is computationally less expensive than BOA, the quality of the model was not as effective as BOA in the case of multi modal problems beyond trap-5. The authors state that, given the efforts needed to train, it is not feasible to be used in a noisy training set. DBM-EDA could not find the global optimum in complex problems when allotted the same population size as BOA.

3.2.3 Autoencoders and Denoising Autoencoder EDA

Autoencoders [9, 10, 37] are neural networks that learn an abstract representation of the input with an encoder layer and then perform a reconstruction step using a decoder to convert the representation to an output that approximates the distribution of training examples. Autoencoders have been used as generative modeling techniques in dimensionality reduction and within feature learning [38].

One of the methods employed to prevent such over-fitting is to use Denoising Autoencoders (DAE) [39]. This approach intentionally corrupts the input or training sample with a certain type of noise (uniform, Gaussian etc.), and the denoising autoencoder is trained to filter out the noise in a denoising layer. Thereafter reconstruction of the output from a "cleaned" input occurs. This makes denoising autoencoders a robust filter to noise within the training samples. Denoising Autoencoder EDA (DAE-EDA) [11] uses denoising autoencoders as generative models for EDA. DAE-EDA was tested on concatenated deceptive traps, NK Landscapes, and HIFF functions. Its performance was compared to BOA and RBM-EDA. DAE-EDA was found to be much faster than BOA and RBM-EDA, however reduced model quality was observed as a consequence.

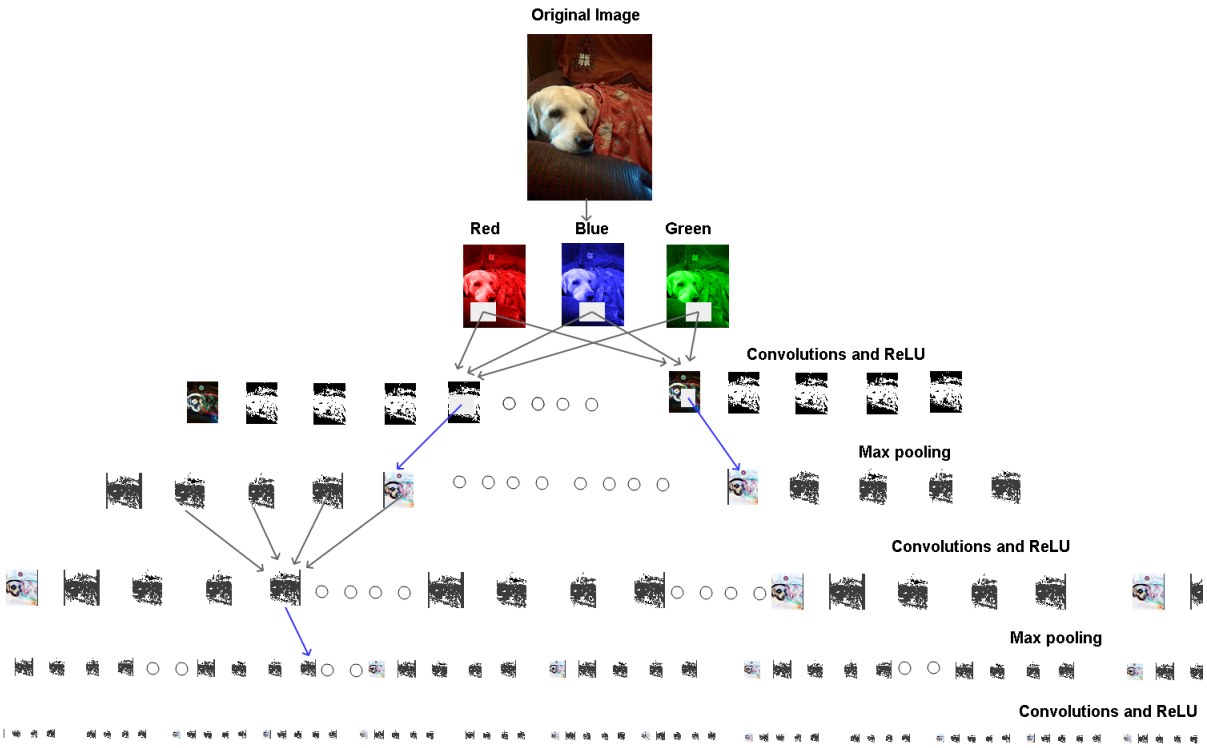


Figure 3.2: Concept of Convolutional Neural Network depitcts an image of a dog broken up into RGB (red, green and blue) inputs before feeding it to the first convolutional layer. Each square box shows a feature map. Image adapted from [18].

3.2.4 Convolutional Neural Networks

A Convolutional Neural Network's (CNN) architecture is comprised of a series of stages (figure 5.4). As described by authors in [18], the initial stages are composed of Convolutional and Pooling layers. The convolutional layer uses filters to execute the convolutional process on the image data provided as input. Each filter in a convolutional layer is designed as a matrix and, in the scope of this paper, we restrict the discussion to 2-dimensional convolutional operations and filters since we are processing 2 dimensional image data only. The filter has a fixed size and scans the image throughout the convolution step, during which it slides in a horizontal direction at a fixed step (known as width of stride), then moves vertically with a fixed step (known as height of stride), before making the next horizontal slide. This process continues

until the entire image has been scanned. At each position on the image where the filter stops, a convolution operation is performed by multiplying a pixel value with a corresponding value in the filter matrix. The values are then added up to get the filter output. The output derived from the filter operation upon the whole input image yields a matrix which is known as feature map. At higher layers, the convolutional operation is applied to the feature maps of the previous layers to generate the feature maps of the current layer. For a given convolutional layer, multiple different filters of standard sizes are utilized to generate maps of the same dimensions (based on the stride values), which forms a set of feature maps. One parameter of a CNN is the number of feature maps available (which corresponds to the number of different filters used per level). A classical trick, known as padding, can be applied at the edge of the image if the filter goes beyond the boundary of a map. In this case, the size of the map is increased by padding zeros on all edges of the map.

The second type of layer used within a CNN is known as Pooling layer, which is used between other CNN layers, and helps to reduce the spatial dimensions of the internal representation. Presence of a pooling layer helps to reduce the number parameters and computational requirements of the network. It also helps to prevent over-fitting of the model to training data. A pooling layer does not change the dimensions of the input image, nor does it contain any parameters as it calculates a fixed function on the input provided. The pooling layer has two hyperparameters: spatial extent of the downsampling filter (known as kernel size) and stride size. Pooling layers can be of max or mean types. For example, consider a *max* pooling layer of kernel size 2x2 and stride of 2. In this case, the downsampling filter will scan a 2x2 region in a slice of the input, and take the max of 4 numbers before moving on to the next 2x2 block (since the stride is 2). This helps to discard 75% of the activations and keeps the overall network structure simple. The mean pooling layer employs the same

process with the exception that it takes average of the values of 2x2 area.

The overall architecture of a CNN involves stacking multiple stages of convolutional layers, reLU, and pooling layers such that it is superseded by more fully connected and convolutional layers. Back propagation of gradients is performed through a CNN to train all the weights. A typical CNN structure used in practical applications can have have more than 100 layers, millions of weights, and connections in the range of billions. In recent years, due to advancements in computing hardware and efficient use of GPUs, CNNs have become extensively used for tasks involving recognition and detection [40–45].

3.3 Variational Autoencoder EDA

3.3.1 Variational Autoencoders

Variational Autoencoders (VAE) [12] [46] improve upon traditional autoencoders in their explorative capabilities by generating data from continuous latent spaces representation, which allows for smooth interpolation and randomized sampling by expanding the latent search space horizon beyond what the training samples represent. While traditional autoencoders are able to introduce minor random variations in the input provided, VAEs allow for a greater degree of control of the direction and magnitude of the variation produced in the output by the hidden model [47]. Autoencoders described in the previous section differ from VAE in that standard autoencoders do not have a probabilistic basis while VAEs allow stochastic generative modeling with tunable probability parameters [48]. We have adapted the following description of VAE from [47].

In [12] and [47], the authors define VAE by considering z as a vector of latent variables in $\mathcal{Z}$ which form a high dimensional space (see Encoder half of figure 5.3).

The z can be sampled using some density function P(z) defined over $\mathcal{Z}$. Equation 5.1 aims at maximizing the probability of X by optimizing θ (X is a data point in our dataset, and θ is a parameter vector in space Θ); where z is sampled with a high probability from P(z).

$$P(X) = Maximize(\int P(X|\mathrm{z};\theta)P(\mathrm{z})dz) \tag{3.1}$$

Hence, in VAE, the objective becomes to model the probabilistic distribution of parameters of the latent space instead of fixed values for them [48]. However, the integral computation is intractable. Some approximations are therefore needed to estimate this distribution. VAE use an approximation $Q(z|X)$, modeled by the encoder, of the posterior distribution $P(z|X)$ and an approximation of $P(X|z)$, modeled by the decoder. Some common choices of the distribution models used in VAEs are Bernoulli Distribution or Multivariate Gaussian distribution. Equation 3.2 gives the probability model for Gaussian distribution having a mean of $f(\mathrm{z};\theta)$, with I being the covariance multiplied by some scalar hyperparameter σ and f representing the decoder network.

$$P(X|\mathrm{z};\theta) = \mathcal{N}(X|f(\mathrm{z};\theta), \sigma^2 * I) \tag{3.2}$$

The values sampled from the model for z need to contain useful information from $P(X|z)$ to produce an output similar to desired target X. Therefore the function $Q(z|X)$ is learned (encoder) in order that the distribution over z can be later used to produce X.

3.3.2 Variational Autoencoder EDA (VAE-EDA)

Authors in [13] employed VAE as a generative modeling technique to be used as the model for EDA (VAE- EDA). The motivation is that VAEs have continuous latent

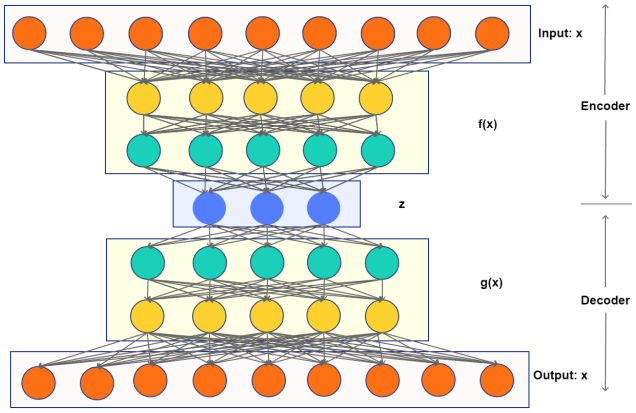


Figure 3.3: Variational Autoencoder structure showing the parent vectors on top layer mapped to higher dimensional spaces in encoder f(x). The z layer provides the probability distribution sampling function, which is sampled by the decoding layers g(x) to yield the output in real dimension at the bottom layer.

spaces and allow for easy interpolation and random sampling. To clarify, if the input space has gaps or discontinuities, standard EDAs have difficulty dealing with the corresponding latent space. With VAE-EDA, a randomized population is initially generated, from which individuals are selected by using Tournament Selection to find and select the subset of fittest individuals. From these fittest individuals, a Gaussian probabilistic VAE model is learned, that is then sampled to derive offspring candidates for the next generation.

3.3.3 Variational Autoencoder EDA with Population Queue (VAE-EDA-Q)

We propose an advanced version of VAE-EDA algorithm in algorithm 1, in which we attempt to control the convergence of our population by using a strategy where the VAE-EDA is updated using a sample built from a sliding window of historical populations. We refer to this algorithm as Variational Autoencoder EDA with Population Queue (VAE-EDA-Q) *. The VAE-EDA-Q model is updated based on the

fittest candidates from the sampled population history, instead of the immediate parent population. We perform tournament selection on candidates from the history to create a new population to update the VAE-EDA-Q model. The latent model of the VAE-EDA-Q is sampled at every time step, and a new population is built by combining the offspring with the fittest candidates in the parent population for the preceding time step. The VAE-EDA-Q is expected to have a greater degree of control over the diversity of the solutions produced. The size of the queue allows us to control this diversity - if the value is too low, the model produces too diverse solutions. And if the value is on the higher side, then the model does not explore as much as desired. The choice of the value for which this parameter is set depends on the difficulty and the nature of the problem to be solved. In our study, the update time step was empirically chosen as 10 to find a value that balances speed of convergence and desirable amount of population diversity for the problems in the scope of this paper.

3.3.4 Variational Autoencoder EDA with Population Queue and Adaptive Variance Scaling (VAE-EDA-Q AVS)

In order to remedy the problem of diversity loss (premature loss of variance in the population), a methodology named Adaptive Variance Scaling (AVS) was suggested in [14] [49] that helped scale the variance beyond the maximum likelihood estimate of the model generated from the parent population. In AVS, a coefficient is used to scale the variance of sampling, which is known as variance multiplier c^{AVS}. If the variance of the model is Σ, then the model is sampled using a variance of $c^{AVS}\Sigma$. If at each iteration we find an improvement in the best fitness, then the c^{AVS} is multiplied by $\eta^{INC} > 1$ to help enlarge the variance and increase the exploration of the search space .

*Source code is available at https://github.com/sourodeep/vaeedaq

Algorithm 1 Variational Autoencoder EDA with Population Queue - VAE-EDA-Q

$i \leftarrow 0$
$QueueSize \leftarrow 10$
$PopulationHistory[] \leftarrow null$ ▷ Define a Queue to store populations from different iterations
Generate Initial Population P(0) ▷ Assume n to be population size
$PopulationHistory[i].PUSH(P(0))$ ▷ Insert initial population into the queue of populations
while Global Optimum not found OR Current Iteration < Max_Iteration **do**
    Select promising solutions N(i) from P(i) using TournamentSelection
    if $i \geq QueueSize$ **then**
        PopulationHistory.POP() ▷ Remove oldest Population from head of the Queue
    end if
    $SampledPopulationHistory \leftarrow$ **TournamentSelection**$(PopulationHistory[])$ ▷ Selecting n candidates
    Build/Update Variational Autoencoder Model M(i) from SampledPopulation-History
    Sample M(i) to produce next generation O(i)
    $P(i+1) \leftarrow N(i) + O(i)$
    $PopulationHistory.PUSH(P(i+1))$ ▷ Insert newly generated population into the queue
    $i \leftarrow i + 1$
end while

If no improvement is observed in fitness values, the variance multiplier is multiplied by a factor of $\eta^{DEC} \in [0, 1]$, which helps in reducing the range of exploration. Generally, in order to maintain symmetry in scaling, $\eta^{INC} = 1/\eta^{DEC}$. We have applied the AVS method to the standard deviation vectors of VAE-EDA-Q at the layer of hidden neurons that model the latent space (z layer in figure 5.3) to dynamically control the amount of exploitation and exploration at the sampling step of each iteration. We will refer to this algorithm as VAE-EDA-Q AVS *. VAE-EDA-Q AVS was able to solve benchmark problems of even greater difficulty than VAE-EDA-Q such as Trap 11 and 13, as well as NK Landscapes of up to epistasis (k) 10.

3.3.5 CNN Architecture Optimization

We used a predefined set of building blocks of CNN to build the population of VAE-EDA-Q AVS, where each candidate in the population represents a combination of building blocks to be used for the construction of a model to be learned for classification of the target image dataset. At the initial iteration, the population is randomly generated using a fixed encoding strategy, which will be discussed subsequently. What follows is a standard VAE-EDA-Q AVS procedure (algorithm 1). The fittest candidates, representing specific CNN architectures, are selected using tournament selection, and new offspring are generated by sampling the latent model generated from the parent population. The next generation of solutions are obtained by combining the fittest parents with the offspring. We will refer to this algorithm as VAE-EDA-Q AVS.

Our encoding strategy follows that of [22] to facilitate a standardized comparison with their results. A CNN is comprised of several layers such as convolutional layers, pooling layers, and fully connected layers. We have used Skip layers, as defined in [22], which uses skip connections directly, thereby replacing the convolutional layer at CNN

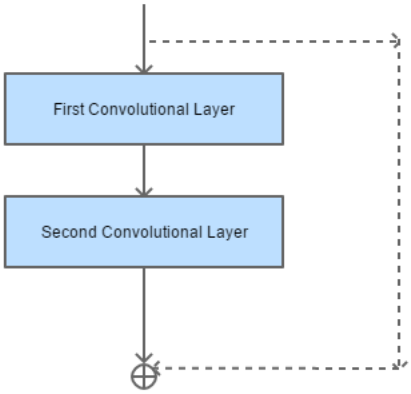


Figure 3.4: Encoding of a Skip layer [22]

creation. A Skip Layer (figure 3.4) uses two convolutional layers and a skip connection that connects the input of the first convolutional layer to the output of the second one. The motivation for this design is drawn from ResNet [50], and the previous research has proven ResNet to be highly effective [51] [52] [23].

The parameters of a convolutional layer includes number of feature maps, stride size, filter size, and type of convolutional operation. In our encoding strategy, we use the same settings for our convolutional layers. The filter size is set to 3 x 3, the stride size is set to 1 x 1, and only *same* convolutional operation is used. Parameters for a skip layer include the number of feature maps in the two constituent convolutional layers (F1 and F2 in algorithm 2). The pooling layer sizes are set to 2 x 2 applied to stride sizes and kernel sizes. The choice of these parameters is inspired from the designs of building blocks of manually created CNNs [51] [52].

Algorithm 2 shows the procedure used to initialize the population for VAE-EDA-Q AVS, where PMax number of initial candidates are generated. Each candidate is stored as a linked list and represents a CNN, where the length S of the candidate

representing the depth of the CNN is uniformly and randomly initialized. After a linked list (LL representing the CNN) is initialized, each node of the linked list is configured either as a pooling layer or a skip layer stochastically with each node having an equal probability of being a skip or a pooling layer. Additionally, we explore the possibility of including two consecutive pooling layers. If a node is configured as a skip layer, the number of filters in the two convolutional layers are uniformly and randomly generated. Conversely, if a node is configured as a pooling layer, the pooling type is set as Max or Mean with equal probability.

Algorithm 2 Population Initialization of VAE-EDA-Q AVS based CNN architecture search

```
    Input: Size of Population (PMax).
    Output: Initial Population P_0.
  P_0 ← NULL
  while |P_0|< PMax do
      S ← RandomInteger                    ▷ Size of Candidate gives Depth of CNN
      LL ← Create Linked List containing S nodes of CNN
      for each Node N in LL do
          r ← Generate Real Number between 0 and 1
          if r < 0.5 then
              N.Type ← Skip_Layer
              N.F1 ← Random Integer (> 0)
              N.F2 ← Random Integer (> 0)
          else
              N.Type ← Pooling_Layer
              q ← Random Number Uniformly Generated (0,1)
              if q < 0.5 then
                  N.P1 ← Max                                   ▷ Pooling Type
              else
                  N.P1 ← Mean                                  ▷ Pooling Type
              end if
          end if
      end for
      P_0 ← P_0 ⋃ LL
  end while
```

Fully connected layers are not used within our encoding strategy - only pooling

layers and skip layers are used. Fully connected layers have not been used in the proposed encoding following [22], as they lead to over-fitting [53] caused by dense connections [54]. Moreover, the number of fully connected layers and neurons within each layer is a parameter that is difficult to optimize. Removing the fully connected layer reduces the search space required to explore and additionally keeps the search tractable.

3.3.6 CNN Model Training and Evaluation

For a given individual in the VAE-EDA-Q AVS population, a CNN is coded and a softmax classifier is added on top it [55]. The classes are determined based on the target image classification dataset. At the decoding step of the CNN, a rectifier activation [56] is used, and then the output is batch normalized [57] as per convention in designing CNNs [50]. Afterwards, the CNN is trained by using Stochastic Gradient Descent (SGD) [58] optimizer on the training data and evaluated on the validation dataset. The best classification accuracy on the validation dataset determines the fitness of the individual in the population of EDA.

3.4 Experiments

3.4.1 Core Experiments on VAE-EDA-Q and VAE-EDA-Q AVS

Test Problems

We have evaluated VAE-EDA-Q and VAE-EDA-Q AVS on concatenated deceptive trap functions of order 5, 7, and 9, as well as NK Landscapes (with the epistasis k varying form 4 to 6), which are considered benchmark test problems [9]. Additionally,

VAE-EDA-Q AVS was also tested on Trap Functions of order 11 and 13 and also on NK Landscape of k = 6 to k = 8. The difficulty of these problems increased linearly with the problem size (given by the number of decision variables) as well as exponentially with the degree of dependencies between parameters (the order of the trap, for example). Therefore, the main driving factor of complexity for these problems is the degree of dependencies.

In concatenated traps [59] of order k (referred to as Trap-K), a solution vector x is divided in subsets of size k, where each subset is used to create a deceptive trap. Within each subset, the variables are dependent upon each other yet independent of the variables within other subsets. Thus, the problem is decomposable into independent sub-problems. Hence, the fitness of each subset contributes to the global fitness as a sum of fitness of individual traps.

Assume p to be an assignment of k values (0 or 1) to an arbitrary subset of variables, then the fitness contribution for that subset is given by Equation 7. The fitness of the trap increases with increasing zeros, unless the assignment are all values of one. The deception is a result of increasing numbers of 1 within a subset causing the fitness to decrease, even though the solution gets closer to global optimum (consisting of an assignment of 1 for all variables). The size of the subset or trap, given by k, increases the complexity of the problem as the value of K increases.

$$f(a) = \begin{cases} k & \text{if } \sum_{i}^{i+k-1} p_i = k \\ k - (\sum p_i + 1) & \text{if } otherwise \end{cases} \tag{3.3}$$

NK Landscapes [60] are functions defined by two parameters: the total number of binary variables in the problem (N) and the size of neighborhood for each variable (K). The fitness function of each bit is provided by a lookup table generated randomly beforehand that determines the fitness of the bit and its neighbors. For example,

consider a problem of N = 4 and K = 1 (the problem contains 4 variables in total), then the fitness of the assignment of all variables f(n) is composed of the sum of the fitness of sub-problems f_i^N, one sub-problem for each variable. The fitness of one variable depends on its value as well as the value of its next k variables (one in our example since K=1).

The total fitness of a solution f(n) in NK Lanscapes is the sum of all fitness components (subsets of a bit or variable and its k neighbors). Formally equation 3.4 [7] shows how the total fitness $f(n)$ is governed by the sum of individual fitness functions f_i^N. f_i^N evaluates fitness of a sub-problem based on the assignment of the first variable x_i as well as its k successors x_{i1} to x_{iK}. Therefore, the value of each decision variable x_i affects the fitness of K+1 sub-functions. Thus, each decision variable assignment impacts multiple fitness components due to this overlap among neighboring variables. These dependencies make it impossible to separate and independently solve the sub-problems.

$$f(x) = 1/N \sum_{i=1}^{N} f_i^N(x_i, x_{i1}, ...x_{iK}) \tag{3.4}$$

In most cases NK Landscapes are NP-Complete for k > 1, which poses a challenging optimization problem [61]. In order to compare VAE-EDA, VAE-EDA-Q, and VAE-EDA-Q AVS with other state-of-the-art algorithms (BM-EDA [7] and DAE-EDA, [11]) we use similar NK Landscapes with known optima as in [7, 11, 61].

Experimental Setup

We have compared VAE-EDA [13], VAE-EDA-Q, VAE-EDA-Q AVS, BOA [31], hBOA [33], BM-EDA [7], CMSS-BOA [34], DBM-EDA [36], and DAE-EDA [11], on Trap

Functions and NK Landscapes of varying complexity. For each algorithm and problem, we have reported results for an average of 30 runs, which were allowed to run until global optimum was found. For consistency of the results, all algorithms used tournament selection (size 2) without replacement (same as reported in experiments for DBM-EDA and DAE-EDA). We have used the results mentioned by the authors in the original papers for BM-EDA and DAE-EDA in our discussions. For Trap 5, we have compared VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, BOA, hBOA, DAE-EDA, DBM-EDA, and CMSS-BOA and reported the execution times in seconds (log scale) as well as the number of unique fitness evaluations. For Trap 7 and 9, we have compared VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, BOA, hBOA, and CMSS-BOA and reported the execution times as well as the total number of fitness evaluations. DAE-EDA and DBM-EDA could not be compared for Trap 7 and 9 as the results for the algorithm were not reported by the authors due to unavailability of their code.

For NK Landscapes (up to K = 6), we compared VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, BOA, hBOA, DBM-EDA, CMSS-BOA, and DAE-EDA and reported the execution times as well as the total number of unique fitness evaluations. Some of the algorithms could not produce results in a comparable time frame or number of generations for the particular test problem, and, hence could not be included in the results. For Traps of 11 and 13 and also for NK Landscape of K = 8 and K = 10, we have compared only VAE-EDA-Q AVS, BOA, and CMSS-BOA as the other algorithms were either not able to solve these problems or the results were not available from other published research. To the best of our knowledge, this paper is the first to present successful results on Trap 11 and 13 and on NK-landscape with K=10. All experiments were performed on a high performance cluster computing platform provided by SHARCNET (https://www.sharcnet.ca) where base compute node con-

figuration was Intel E5-2683 v4 (Broadwell) 2.1 GHz with 32 cores, 128 GB of memory. For Trap 5 and NK Landscape ($K < 8$), we have used results reported in previously published research. For all other test problems, the tests were performed on the same system configuration.

Implementation

For both algorithms proposed in this paper, VAE-EDA-Q *and VAE-EDA-Q AVS, Nesterov Adam optimizer (NADAM) was used as the optimizer [62]. This provides a better computational performance than stochastic gradient descent. The training of the model used a mini-batch size of 64 samples, and the number of epochs was set as approximately P/64, where P is the size of the training population. The intermediate layers of hidden neurons contained n/2 nodes where n represented the problem size, while the latent dimension hidden layer always contained 20 fixed hidden nodes for all of our experiments. A learning rate of 0.02 was used for both algorithms. For VAE-EDA, we have used the same model architecture as suggested by the authors in [13]. The algorithms were implemented in Python programming language using Keras Deep learning library (https://keras.io/).

3.4.2 Experiments on VAE-EDA-Q AVS based CNN architecture Search

Image Classification Datasets

We have used CIFAR10 and CIFAR100 datasets [30] for image classification in our experiments as they are difficult to solve due to their image size, higher number of image classes (10 or 100), and the presence of rotation, translation (position of object

*Source code is available at https://github.com/sourodeep/vaeedaq

to be classified varies in the image), and noise in the images. They are considered benchmark datasets to gauge the performance of deep learning based image classifier algorithms. CIFAR10 contains ten classes of objects commonly seen within our environment such as cats, dogs, airplanes, automobiles, etc. It consists of 60000 size 32 x 32 RGB images, from which 50000 images are used for training and the rest for testing. CIFAR100 is a similar dataset, with the exception of containing 100 classes with 600 images in each class.

We have split the training dataset into two parts whereby the first 90% is used for training of the CNN and the remaining of the 10% images are used for validation of the model trained. A pre-processing is performed to each image before training in which four zero pixels are padded to every direction of a given image. A fixed sized crop is then sampled randomly and uniformly from the image, and a cropped image can be horizontally flipped with a probability of 0.5. This pre-processing procedure standardizes the images and is used widely in state-of-the-art algorithms [50] [51].

Comparison with State-of-the-art Algorithms

We have compared VAE-EDA-Q AVS to manually designed CNNs, which are known to be highly accurate in classfications [63] such as DenseNet [51], ResNet [50], Maxout [64], VGG [45], Network in Network [65], All-CNN [66], and Highway Network [67]. For ResNet we have used ResNet models of depth 101 and 1202 only.

Our proposed algorithm has been compared with some semi-automatic algorithms that require some form of manual designing before an architecture is discovered. The algorithms used for comparison were Hierarchical Evolution [24], Genetic CNN [21], Block-QNN-S [29], and EAS [28]. EAS is considered semi-automatic because it works on CNN blocks, which have already proven to have good accuracy for the dataset being classified. QNN needs a predefined large CNN, on which the discovered CNNs are

added. Some fully automatic CNN search algorithms have also been compared to our method such as CNN-GA [22], Large-Scale Evolution [23], NAS [26], CGP-CNN [25], and MetaQNN [27]. These algorithms do not require any human intervention in the process.

3.5 Results and Discussion

3.5.1 Core Experiments on VAE-EDA-Q and VAE-EDA-Q AVS

All results presented correspond to an average of 30 independent runs for each method. Figure 5 shows the execution time comparisons between BOA, hBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DBM-EDA, and DAE EDA for Trap 5 problems of problem sizes 25, 50, 75, and 100. We can see that both VAE-EDA-Q and VAE-EDA-Q AVS have better computational performance than VAE-EDA, BOA, and CMSS BOA. While VAE-EDA-Q and VAE-EDA-Q AVS outperform DBM-EDA, VAE-EDA is unable to do so. The execution of time of DAE-EDA, as reported by the authors, turns out to be slightly lower than those of VAE-EDA-Q and VAE-EDA-Q AVS.

For the number of fitness evaluations (figure 5.6), it was observed that DAE-EDA performs more fitness evaluations than all other algorithms. VAE-EDA-Q and VAE-EDA-Q AVS perform less fitness evaluations than VAE-EDA, BOA, and CMSS-BOA, with BOA being the highest. We also find that VAE-EDA-Q has better performance than VAE-EDA both in terms of execution times and fitness evaluations. The number of performance evaluations of DBM-EDA lies between that of VAE-EDA and VAE-EDA-Q.

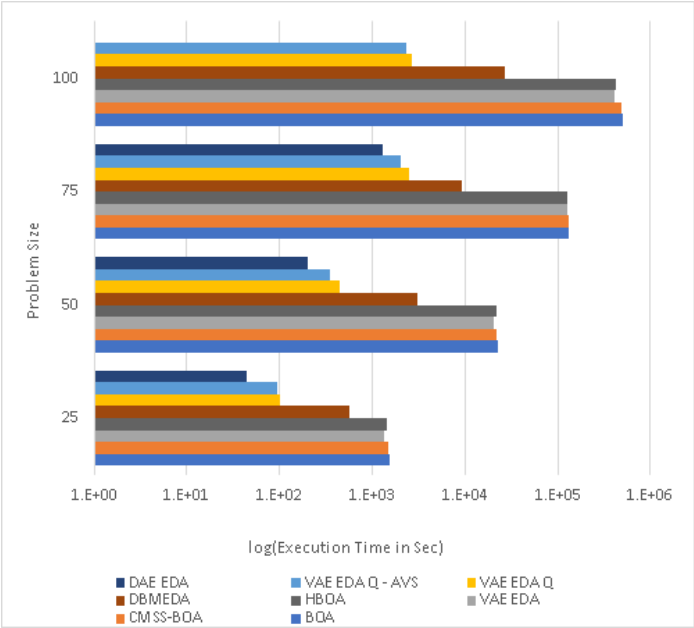


Figure 3.5: Execution time for Trap 5 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DBM-EDA, and DAE-EDA. VAE-EDA-Q and VAE-EDA-Q AVS have better performance than VAE-EDA, BOA, and CMSS-BOA.

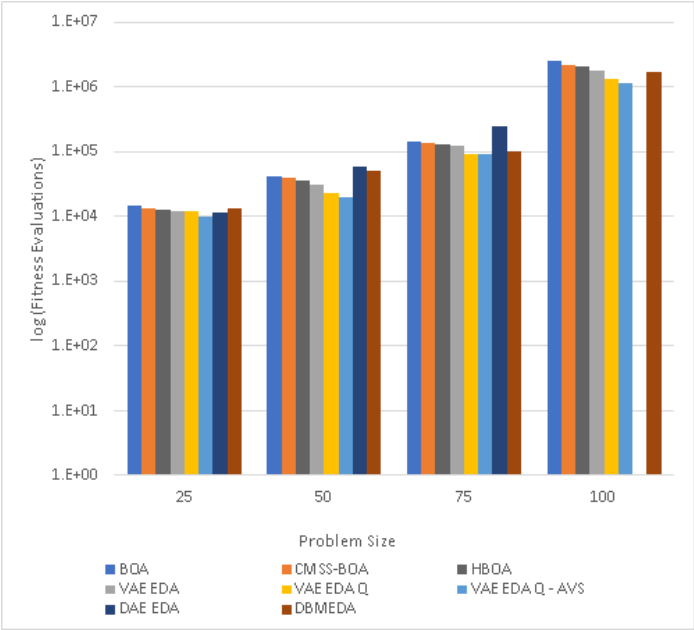


Figure 3.6: Number of fitness Evaluations for Trap 5 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DBM-EDA, and DAE-EDA. VAE-EDA-Q has the lowest number of fitness evaluations compared to all the other algorithms

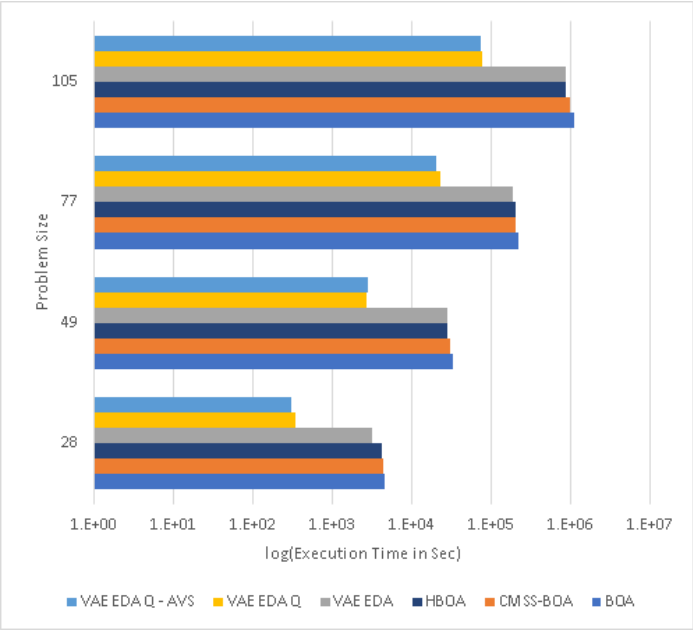


Figure 3.7: Execution time for Trap 7 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q AVS, and VAE-EDA-Q. VAE-EDA and VAE-EDA-Q perform better than BOA and CMSS-BOA, with VAE-EDA-Q having the lowest execution time

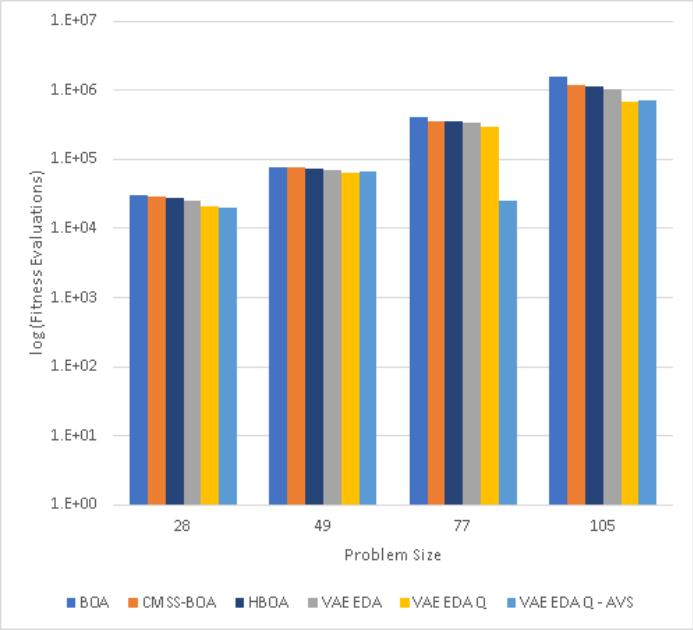


Figure 3.8: Number of fitness Evaluations for Trap 7 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q AVS, and VAE-EDA-Q. VAE-EDA and VAE-EDA-Q perform better than BOA and CMSS-BOA, with VAE-EDA-Q having the lowest number of fitness evaluations

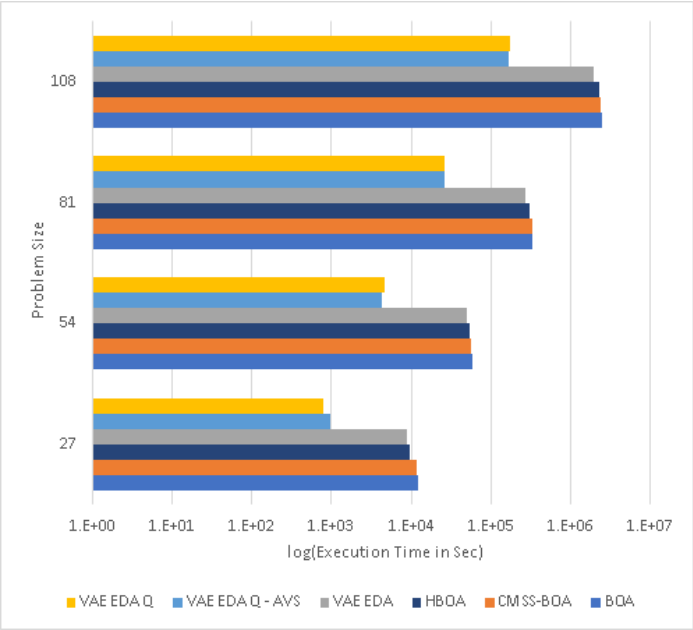


Figure 3.9: Execution time for Trap 9 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q AVS, and VAE-EDA-Q. VAE-EDA-Q AVS and VAE-EDA-Q perform better than VAE-EDA, BOA and CMSS-BOA, with VAE-EDA-Q having the lowest execution time

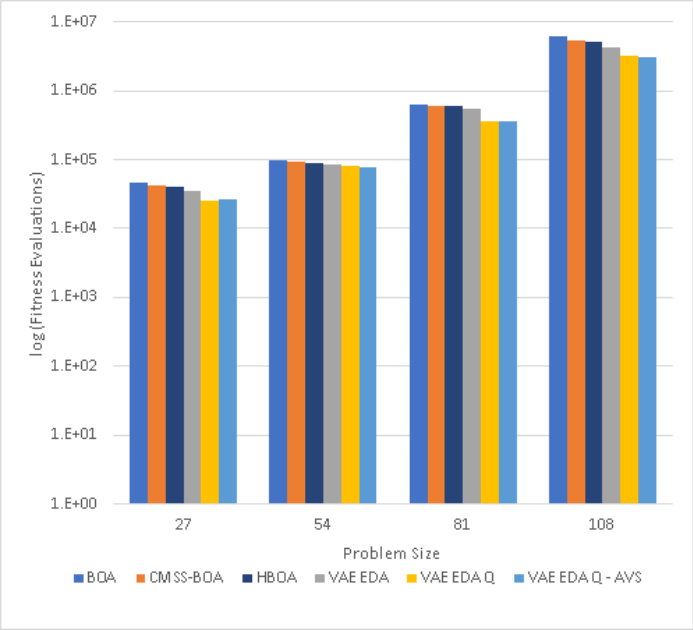


Figure 3.10: Number of fitness Evaluations for Trap 9 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q AVS, and VAE-EDA-Q. VAE-EDA-Q and VAE-EDA-Q AVS perform better than VAE-EDA, BOA and CMSS-BOA, with VAE-EDA-Q having the lowest number of fitness evaluations

Figures 5.7, 5.8, 5.9, and 5.10 shows similar results for VAE-EDA, VAE-EDA-Q AVS, VAE-EDA-Q, CMSS-BOA, hBOA, and BOA for higher order traps of order 7 and 9 respectively. For trap 7 and trap 9, DAE-EDA and DBM-EDA could not be compared as the result for higher order traps was not provided by the authors in their original research. We find that VAE-EDA-Q and VAE-EDA-Q AVS have approximately one-tenth the execution time requirement of BOA and perform approximately half the number of fitness evaluations as BOA.

In addition to having a better performance than other algorithms tested, VAE-EDA-Q AVS was able to solve traps of orders 11 and 13 for all 30 runs with approximately 20 to 50 times less computational time requirement than BOA and CMSS-BOA as shown in the graphs in figures 3.11-3.14. VAE-EDA and VAE-EDA-Q could not solve these problems, and no results were reported for DAE-EDA and DBM-EDA for these problems. Moreover, the success rate of BOA, hBOA, and CMSS-BOA was only about 23% (approximately 7 out of 30 runs were able to find the global optimum), while the success rate of VAE-EDA-Q AVS was 100%. These results prove that VAE-EDA-Q AVS is more capable at successfully exploring the latent space by dynamically varying the degree of exploration based on fitness.

In order to test the statistical significance of the results, we performed Kruskal Wallis ANOVA [68] separately for the execution time and the number of fitness evaluations independently for each trap order, and all the differences turned out to be significant (p-value < 0.05). For Trap 5 specifically, we performed an additional Kruskal Wallis ANOVA test on DAE-EDA, VAE-EDA-Q, and VAE-EDA-Q AVS for execution time, and the result shows that the differences were not statistically significant (p-value $\bar{0}.149$ for execution time with alpha at 0.05).

Figure 3.15 shows the comparison of BOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DAE-EDA, and DBM-EDA on NK Landscapes with problem sizes

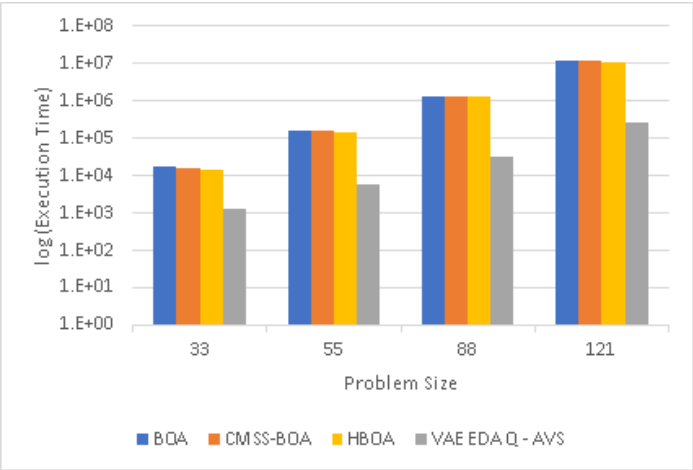


Figure 3.11: Execution time for Trap 11 comparing BOA, HBOA, CMSS-BOA, and VAE-EDA-Q AVS. VAE-EDA-Q AVS has lower execution time compared to BOA and CMSS-BOA for all problem sizes.

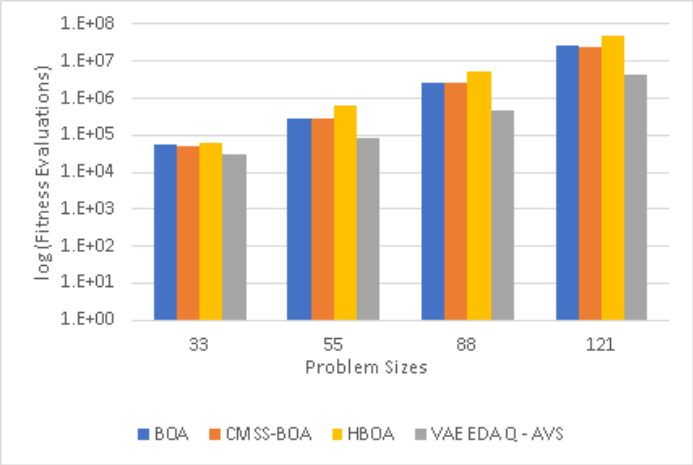


Figure 3.12: Number of fitness Evaluations for Trap 11 comparing BOA, HBOA, CMSS-BOA, and VAE-EDA-Q AVS. VAE-EDA-Q AVS has lower number of fitness evaluations compared to BOA and CMSS-BOA for all problem sizes.

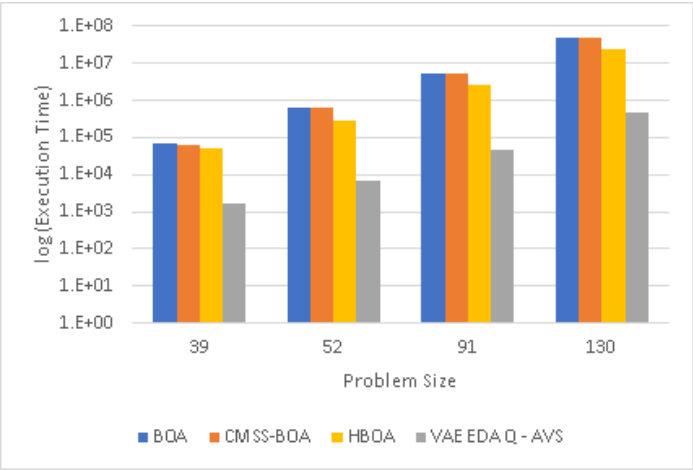


Figure 3.13: Execution time for Trap 13 comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q AVS, and VAE-EDA-Q. VAE-EDA-Q AVS has lower execution time compared to BOA and CMSS-BOA for all problem sizes.

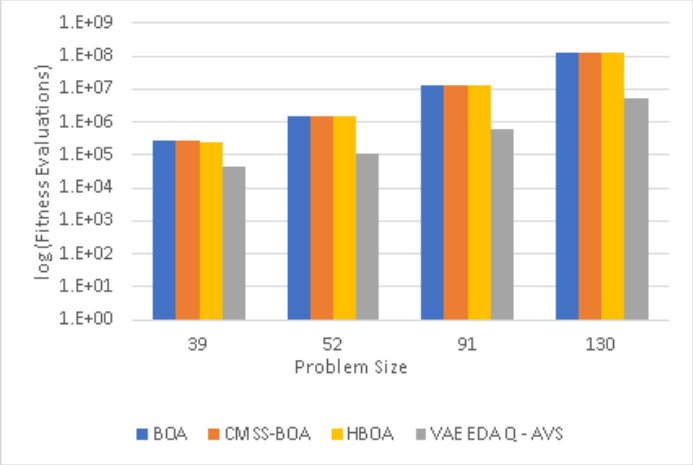


Figure 3.14: Number of fitness Evaluations for Trap 13 comparing BOA, HBOA, CMSS-BOA, and VAE-EDA-Q AVS. VAE-EDA-Q AVS has lower number of fitness evaluations compared to BOA and CMSS-BOA for all problem sizes.

30 and 34 and neighborhood size varying from 4 to 10. VAE-EDA-Q was found to have a slightly higher execution time than DAE-EDA, though the differences were not significant according to Kruskal Wallis ANOVA test. VAE-EDA-Q was also found to perform the smallest number of fitness evaluations (half or less) compared to all other algorithms tested for most cases.

While no results were presented for DBM-EDA and DAE-EDA for neighborhood size of 6 and above, we have performed the experiments to compare VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, BOA, and CMSS-BOA on the same problem sizes but with k varying between 6 and 10. It was observed that VAE-EDA-Q and VAE-EDA-Q AVS have better performance than VAE-EDA-Q, BOA, and CMSS-BOA in terms of execution time as well as number of fitness evaluations.

Moreover, VAE-EDA-Q AVS was able to solve NK Landscapes with neighborhood size 8 and 10 at a better success rate than BOA. For BOA, the success rate was 40%, while VAE-EDA-Q AVS had a success rate of 100%. The execution time and the number of fitness evaluations for VAE-EDA-Q AVS was also better than BOA for these problems. VAE-EDA-Q AVS was 10 - 30 times faster than BOA for NK Landscape problems. The other algorithms were either not able to solve these problems or no results were reported in past research for these algorithms for problems of this complexity. The differences in computational time observed in these comparisons were statistically significant (Kruskal Wallis ANOVA test with p-values < 0.05).

VAE-EDA- Q and VAE-EDA- Q AVS were found to have very similar execution time as compared to DAE-EDA without any statistically significant differences for trap problems and NK Landscapes. VAE-EDA-Q and VAE-EDA-Q AVS were found to outperform all other algorithms including DAE-EDA in number of fitness evaluations. This can be attributed to the fact that BOA, CMSS-BOA, DBM-EDA, and DAE-EDA require larger population sizes to find global optimum and that the VAE

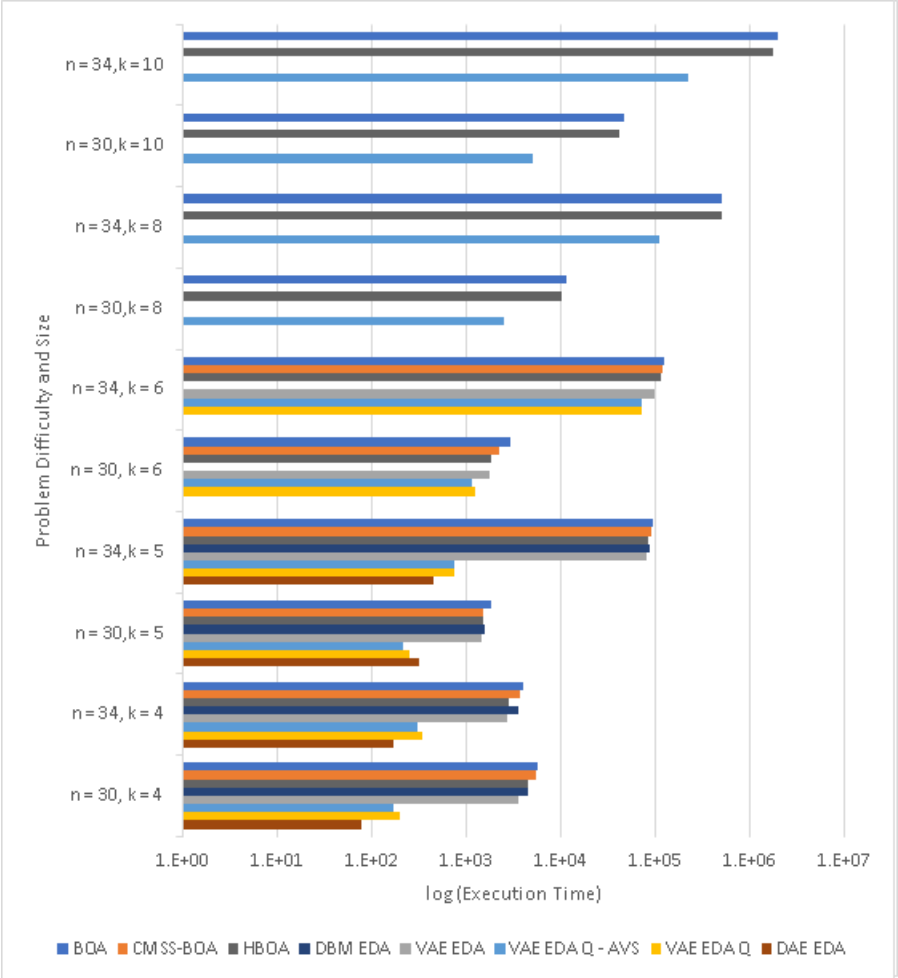


Figure 3.15: Execution time for NK Landscapes comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DBM-EDA, and DAE-EDA. DAE-EDA has lowest execution time upto k = 5. VAE-EDA-Q AVS is able to solve K = 8 and K = 10 at better performance than BOA.

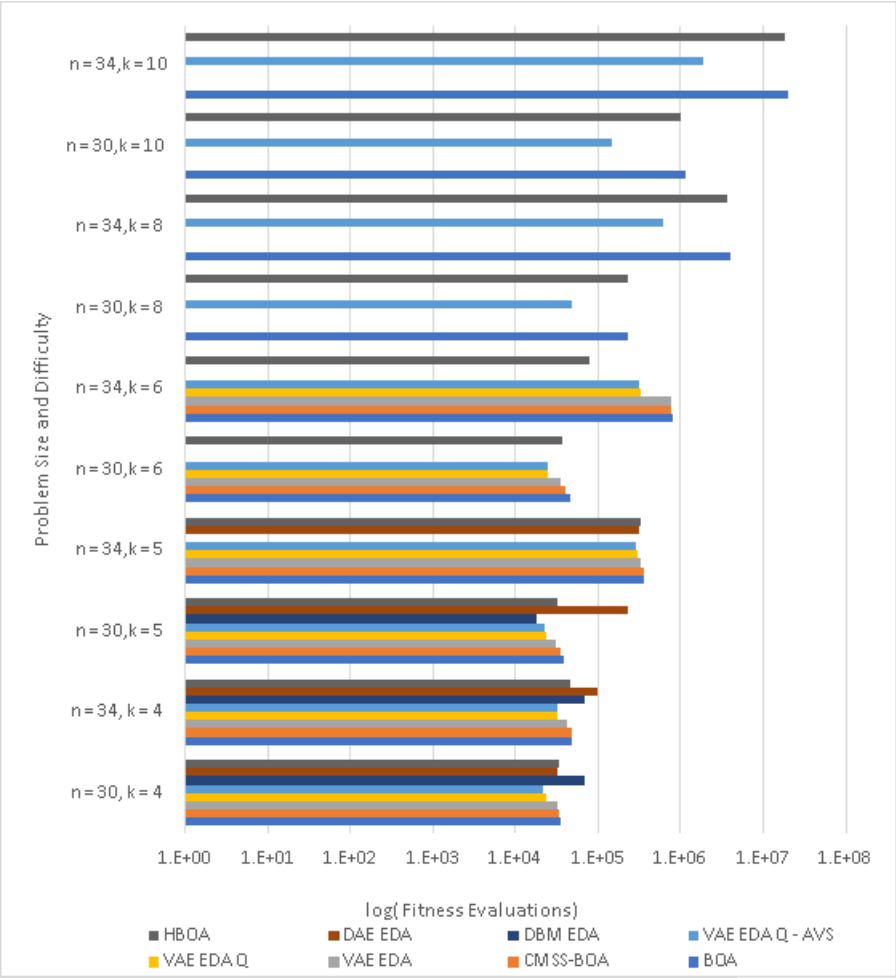


Figure 3.16: Number of fitness evaluations for NK Landscapes comparing BOA, HBOA, CMSS-BOA, VAE-EDA, VAE-EDA-Q, VAE-EDA-Q AVS, DBM-EDA, and DAE-EDA. VAE-EDA-Q has lowest number of fitness evaluations compared to all other algorithms. VAE-EDA-Q AVS is able to solve K = 8 and K = 10 at better performance than BOA

is particularly efficient to detect and model the main dependencies. For BOA and CMSS-BOA, the necessary population size is calculated as $T_p = 2^k \times n^{1.05}$ [32] [69] where k is the size of the trap and n is the problem size. Hence, with increasing problem difficulty (order of the trap or k), the necessary population size increases exponentially. Conversely, VAE-EDA-Q and VAE-EDA-Q AVS are better at exploring latent spaces using significantly smaller population sizes. The maximum population size required to solve trap 9 for a problem of size 108 was 2400 for VAE-EDA and VAE-EDA-Q. Conversely, for DAE-EDA and DBM-EDA, the population size requirement approached 16000 - even for trap 5 problems. While for BOA and CMSS-BOA, the population size requirement was 69000 (for trap 9 problem of size 108).

We also observed that VAE-EDA-Q AVS was able to solve problems of greater difficulty than VAE-EDA, VAE-EDA-Q, DAE-EDA, and DBM-EDA. These problems included Trap 11, Trap 13, and NK Landscapes with neighborhood size 8 and 10. While BOA, hBOA, and CMSS-BOA were also able to solve some of these problems, their success rate was 40% or less. This is in contrast to VAE-EDA-Q AVS, which had a success rate of 100%.

3.5.2 Experiments on VAE-EDA-Q AVS based CNN Architecture Search

We have compared VAE-EDA-Q AVS with other state-of-the-art algorithms based on two metrics - the classification accuracy and the number of parameters shown in Table 3.1. Classification accuracy represents the general success of the algorithm when presented a task to classify an image. The number of parameters represents the complexity of the CNN discovered by the generative model or static number of parameters, in the case of manually designed CNNs. We have used publicly reported results for state-of-the-art algorithms when comparing the results of our proposed

Table 3.1: Comparison between VAE-EDA-Q AVS and other state-of-the-art algorithms based on percentage of classification accuracy and number of parameters of the Convolutional Neural Network discovered

Group	**Algorithms**	**CIFAR10 Accuracy**	**CIFAR100 Accuracy**	**Number of Parameters**
Manually created CNNs	ResNet (depth = 101)	93.57	74.84	1.7M
	ResNet (depth = 1,202)	92.07	72.18	10.2M
	DenseNet	94.17	76.58	27.2M
	VGG	93.34	71.95	20.04M
	Maxout	90.70	61.40	
	Network in Network	91.19	64.32	
	Highway Network	92.40	67.66	
	All-CNN	92.75	66.29	1.3M
Semi-Automatic	Genetic CNN	92.90	70.97	
	Hierarchical Evolution	96.37		
	EAS	95.77		23.4M
	Block-QNN-S	95.62	79.35	6.1M
Fully Automated	Large Scale Evolution	94.60		5.4M
	Large Scale Evolution		77.00	40.4M
	CGP-CNN	94.02		1.68M
	NAS	93.99		2.5M
	Meta-QNN	93.08	72.86	
	CNN-GA	95.22		2.9M
	CNN-GA		77.97	4.1M
	VAE-EDA-Q AVS CNN	95.03		2.1M
	VAE-EDA-Q AVS CNN		75.86	3.6M

methodology. An empty cell indicates that the parameter is not applicable or that the result was not reported in original research. For VAE-EDA-Q AVS, we have taken the final CNN discovered by the model and trained it over 30 independent runs to find the average result reported in Table 3.1.

Comparison with Manually Created CNNs

For CIFAR 10, we observe an increase in classification accuracy of 2.15%, 0.4%, and 1.3% for VAE-EDA-Q AVS when compared to ResNet (depth 1202), DenseNet, and VGG respectively. VAE-EDA-Q AVS uses 4.7, 12.5, and 10 times less parameters than ResNet (depth 1202), DenseNet, and VGG respectively, indicating that VAE-EDA-Q AVS discovers simpler networks that have better classification accuracy.

While the number of parameters in VAE-EDA-Q AVS is higher for CIFAR10 compared to ResNet (depth 101) and All-CNN, the classification accuracy is 1.05% and 3% higher respectively. Moreover, for CIFAR10, VAE-EDA-Q AVS has a higher classification accuracy than Maxout, Network in Network, and Highway Network. We observe similar improvement in classification accuracy for CIFAR100, where VAE-EDA-Q AVS outperforms all reported results in the manually created CNN category while using fewer parameters than most of the algorithms within the first group. VAE-EDA-Q AVS uses more parameters for CIFAR100 than ResNet (depth 101) and All-CNN, but yields better accuracy.

Comparison with Semi-Automatic Algorithms

In the case of the Semi-Automatic group of algorithms for CIFAR10 and CIFAR100, VAE-EDA-Q AVS has 1.79% and 6.89% times higher accuracy, respectively, compared to Genetic CNN. For the other algorithms in this group, we find decreased accuracy for VAE-EDA-Q AVS. However, we also find that VAE-EDA-Q AVS uses a significantly

less number of features in most cases. Still, it should be mentioned that VAE-EDA-Q AVS does not require any domain expertise, while all algorithms in this category require human expertise in some form. For CIFAR10, VAE-EDA-Q AVS has 0.986 times the accuracy produced by Hierarchical Evolution, and 0.992 times the accuracy obtained from EAS. VAE-EDA-Q AVS uses 11.14 times less parameters in EAS. On comparing Block-QNN-S to VAE-EDA-Q AVS for CIFAR10 and CIFAR100, we find that VAE-EDA-Q AVS has 1.09% and 4.39% lower accuracy, respectively. However, we also find that VAE-EDA-Q AVS uses 2.9 and 1.7 times less parameters as compared to Block-QNN-S for CIFAR10 and CIFAR100 respectively.

In order to solve real world problems, algorithms in this group require extensive human support. For example, a preliminary requirement of EAS is to provide a manually designed and tuned CNN to begin with, which the algorithm can fine tune. If the initially tuned network architecture is not well designed, then the algorithm's performance will degrade. Also in cases of Block-QNN-S and Hierarchical Evolution, the output CNN given by the algorithms cannot be applied directly to solve the task. They must be augmented into a large scale CNN that require manual design in order for the accuracy to be at par. If the larger network is not manually designed well, then the performance of these algorithms degrade significantly. VAE-EDA-Q AVS, on the other hand, does not have any requirement of manual intervention and can be readily used on any real world problem

Comparison with Fully Automatic Algorithms

Finally, for the fully automated group of algorithms, and for CIFAR10 and CIFAR100, we find that VAE-EDA-Q AVS has 0.45% higher and 1.48% lower accuracy compared to Large Scale Evolution, at 2.57 and 11.22 times less parameters than Large Scale Evolution.

We observe a slight increase in accuracy of VAE-EDA-Q AVS on CIFAR10 at 1.07%, 1.1% and 2.09% higher compared to CGP-CNN, NAS, and Meta-QNN respectively. However, CGP-CNN uses 0.8 times more parameters used by VAE-EDA-Q AVS. VAE-EDA-Q AVS uses 1.19 times less parameters compared to NAS. For Meta-QNN on CIFAR100, we observe a 4.11% increase in accuracy for VAE-EDA-Q AVS.

Finally, for CNN-GA, on CIFAR10 and CIFAR100 datasets, we observe a 0.19% and 2.7% decrease in accuracy in VAE-EDA-Q AVS respectively. This decrease in accuracy is offset by an increase in performance where VAE-EDA-Q AVS uses 1.38 and 1.14 times less parameters than CNN-GA for CIFAR10 and CIFAR100 respectively.

An overview of the results suggest that VAE-EDA-Q AVS performs better than manually created CNNs in terms of classification accuracy as well as number of parameters (complexity of the CNN) in almost all cases. In the case of semi-automatic algorithms, VAE-EDA-Q AVS has a very close accuracy, and sometimes better accuracy, at far less network complexity. For the fully automatic group, we observe an increase in accuracy and a decrease in network complexity in almost every case except CNN-GA. For CNN-GA, we offer a trade-off for accuracy at an increased speed of execution with a simpler network architecture. It should also be noted that Manual CNNs and Semi-Automatic CNNs require human expertise, whereas fully automated algorithms like VAE-EDA-Q AVS can be applied to any real world problem without any human intervention.

3.6 Conclusion

We have presented two novel approaches for Estimation of Distribution algorithms based on Variational Autoencoders for exploring the use of latent variable modeling in EDAs. The first approach uses a variational autoencoder to model the population

by building its model based on a queue of populations that update at each iteration (VAE-EDA-Q). The second algorithm uses Adaptive Variance Scaling (AVS) [14] along with Queue based Variational Autoencoder model (VAE-EDA-Q AVS) to dynamically adjust the variance while sampling the latent space based upon observance of improvement in the fitness generated. VAE-EDA-Q and VAE-EDA-Q AVS were compared to the classic Bayesian network EDA algorithms (such as BOA, hBOA, and CMSS-BOA) as well as state-of-the-art algorithms (such as the standard VAE-EDA, DAE-EDA, and DBM-EDA). The algorithms were tested on benchmark problems composed of Trap problems and NK Landscapes.

The results obtained were very promising in that VAE-EDA-Q required the lowest number of fitness evaluations compared to all other algorithms for all of the problems tested. In terms of execution time, VAE-EDA-Q had a slightly higher execution time compared to DAE-EDA (with implementation in Octave) for trap 5 and NK Landscape with neighborhood size of 5 problems, though the differences were not statistically significant. This can be attributed to differences in implementation and room for code optimization since the number of fitness evaluations performed by VAE-EDA-Q was lower than DAE-EDA. Moreover, and more importantly, the problems used for the comparisons utilized fitness function that are very fast to compute. For real life problems, where the fitness function can be very computationally expensive to compute, the drastic reduction in the number of fitness evaluations of VAE-EDA-Q and VAE-EDA-Q AVS would lead to an important reduction in the overall computational time. In addition to having a better performance than other algorithms, VAE-EDA-Q AVS is able to solve problems much more complex than the ones that can be solved by VAE-EDA, VAE-EDA-Q, DBM-EDA, and DAE-EDA. While BOA, hBOA, and CMSS-BOA were also able to solve some of these problems, the success rate was much lower and the number of fitness evaluation was much greater (10 to 30

times more) than displayed by VAE-EDA-Q AVS.

The success of VAE-EDA-Q and VAE-EDA-Q AVS is due to their use of generative modeling and their ability to find solutions using about 12% of the population size required by state-of-the-art algorithms (such as DAE-EDA and DBM-EDA) and 3% of the population size required by Bayesian modeling algorithms (such as BOA, hBOA and CMSS-BOA) respectively. We have also presented results for problems of greater difficulty than the ones presented for DBM-EDA and DAE-EDA (such as Trap 7, Trap 9, Trap 10, Trap 13 and NK Landscapes with K = 6 to K = 10), which demonstrates the robustness of our approaches with scaling complexity. We have demonstrated that VAE-EDA-Q and VAE-EDA-Q AVS outperform VAE-EDA, BOA, hBOA, and CMSS-BOA in execution time and fitness evaluations for both Trap problems and NK Landscapes.

In this paper, we have also proposed an estimation of distribution based automated architecture search algorithm for convolutional neural networks named VAE-EDA-Q AVS. The core algorithm of VAE-EDA-Q AVS was extended to encode arbitrary CNN architectures of dynamically generated depths and to produce optimized offspring candidates of similarly varied CNN architectures. VAE-EDA-Q AVS was tested on CIFAR10 and CIFAR100 benchmark datasets successfully and then compared to various state-of-the-art CNN algorithms. It was observed that VAE-EDA-Q AVS generates CNN models that have 1.5% higher accuracy while requiring 25% less parameters for CIFAR10 on average compared to all other state-of-the-art algorithms, and 6% higher accuracy with 10% less parameters for CIFAR100 on average. This indicates that VAE-EDA-Q AVS is able to discover CNN architectures that are simpler in design and yet provides a better classification accuracy.

Future work involves improving upon VAE-EDA-Q AVS by dynamically changing population queue size based upon the state of the iteration as well as balancing the

amount of exploration and exploitation performed as the algorithm progresses. We expect the new algorithm to have lower execution time and, additionally, a lower number of fitness evaluations while being able to tackle even more complex problems.

Bibliography

[1] S. Baluja, "Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning," Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, Tech. Rep., 1994.

[2] H. Mühlenbein and G. Paass, "From recombination of genes to the estimation of distributions i. binary parameters," in *International conference on parallel problem solving from nature*. Springer, 1996, pp. 178–187.

[3] P. Larrañaga and J. A. Lozano, *Estimation of distribution algorithms: A new tool for evolutionary computation*. Springer Science & Business Media, 2001, vol. 2.

[4] M. Pelikan, D. E. Goldberg, and F. G. Lobo, "A survey of optimization by building and using probabilistic models," *Computational optimization and applications*, vol. 21, no. 1, pp. 5–20, 2002.

[5] M. Hauschild and M. Pelikan, "An introduction and survey of estimation of distribution algorithms," *Swarm and evolutionary computation*, vol. 1, no. 3, pp. 111–128, 2011.

[6] P. Smolensky, "Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter information processing in dynamical systems: Foundations of harmony theory," *MIT Press, Cambridge, MA, USA*, vol. 15, p. 18, 1986.

[7] M. Probst, F. Rothlauf, and J. Grahl, "Scalability of using restricted boltzmann machines for combinatorial optimization," *European Journal of Operational Research*, vol. 256, no. 2, pp. 368–383, 2017.

[8] M. Probst and F. Rothlauf, "Deep boltzmann machines in estimation of distribution algorithms for combinatorial optimization," *arXiv preprint arXiv:1509.06535*, 2015.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[11] M. Probst, "Denoising autoencoders for fast combinatorial black box optimization," in *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, 2015, pp. 1459–1460.

[12] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[13] U. Garciarena, R. Santana, and A. Mendiburu, "Expanding variational autoencoders for learning and exploiting latent representations in search distributions," in *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 2018, pp. 849–856.

[14] J. Grahl, P. A. Bosman, and F. Rothlauf, "The correlation-triggered adaptive variance scaling idea," in *Proceedings of the 8th annual conference on Genetic and evolutionary computation*. ACM, 2006, pp. 397–404.

[15] I. Guyon, K. Bennett, G. Cawley, H. J. Escalante, S. Escalera, T. K. Ho, N. Macia, B. Ray, M. Saeed, A. Statnikov *et al.*, "Design of the 2015 chalearn automl challenge," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[16] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.

[17] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[19] T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.

[20] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 135.

[21] L. Xie and A. Yuille, "Genetic cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1379–1388.

[22] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Automatically designing cnn architectures using genetic algorithm for image classification," *arXiv preprint arXiv:1808.03818*, 2018.

[23] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2902–2911.

[24] H. Liu, K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu, "Hierarchical representations for efficient architecture search," *arXiv preprint arXiv:1711.00436*, 2017.

[25] M. Suganuma, S. Shirakawa, and T. Nagao, "A genetic programming approach to designing convolutional neural network architectures," in *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 2017, pp. 497–504.

[26] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[27] B. Baker, O. Gupta, N. Naik, and R. Raskar, "Designing neural network architectures using reinforcement learning," *arXiv preprint arXiv:1611.02167*, 2016.

[28] H. Cai, T. Chen, W. Zhang, Y. Yu, and J. Wang, "Efficient architecture search by network transformation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[29] Z. Zhong, J. Yan, and C.-L. Liu, "Practical network blocks design with q-learning," *arXiv preprint arXiv:1708.05552*, vol. 1, no. 2, p. 5, 2017.

[30] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Citeseer, Tech. Rep., 2009.

[31] M. Pelikan, D. E. Goldberg, and E. Cantu-Paz, "Linkage problem, distribution estimation, and bayesian networks," *Evolutionary computation*, vol. 8, no. 3, pp. 311–340, 2000.

[32] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "Boa: The bayesian optimization algorithm," in *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*. Morgan Kaufmann Publishers Inc., 1999, pp. 525–532.

[33] M. Pelikan, D. E. Goldberg, and S. Tsutsui, "Hierarchical bayesian optimization algorithm: toward a new generation of evolutionary algorithms," in *SICE 2003 Annual Conference*, vol. 3. IEEE, 2003, pp. 2738–2743.

[34] E. Salehi and R. Gras, "Efficient eda for large opimization problems via constraining the search space of models," in *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*. ACM, 2011, pp. 73–74.

[35] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm," *Machine learning*, vol. 65, no. 1, pp. 31–78, 2006.

[36] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, D. van Dyk and M. Welling, Eds., vol. 5. Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR, 16–18 Apr 2009, pp. 448–455.

[37] Y.-l. Boureau, Y. L. Cun *et al.*, "Sparse feature learning for deep belief networks," in *Advances in neural information processing systems*, 2008, pp. 1185–1192.

[38] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[39] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.

[40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[41] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[42] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[43] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[44] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[46] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," *arXiv preprint arXiv:1401.4082*, 2014.

[47] C. Doersch, "Tutorial on variational autoencoders," *arXiv preprint arXiv:1606.05908*, 2016.

[48] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, pp. 1–18, 2015.

[49] P. A. Bosman, J. Grahl, and D. Thierens, "Enhancing the performance of maximum–likelihood gaussian edas using anticipated mean shift," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2008, pp. 133–143.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[53] D. M. Hawkins, "The problem of overfitting," *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1–12, 2004.

[54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[55] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of electronic imaging*, vol. 16, no. 4, p. 049901, 2007.

[56] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.

[57] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," in *Advances in Neural Information Processing Systems*, 2017, pp. 1945–1953.

[58] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 421–436.

[59] K. Deb and D. E. Goldberg, "Analyzing deception in trap functions," in *Foundations of genetic algorithms*. Elsevier, 1993, vol. 2, pp. 93–108.

[60] S. A. Kauffman and E. D. Weinberger, "The nk model of rugged fitness landscapes and its application to maturation of the immune response," *Journal of theoretical biology*, vol. 141, no. 2, pp. 211–245, 1989.

[61] M. Pelikan, "Analysis of estimation of distribution algorithms and genetic algorithms on nk landscapes," in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. ACM, 2008, pp. 1033–1040.

[62] T. Dozat, "Incorporating nesterov momentum into adam," 2016.

[63] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[64] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," *arXiv preprint arXiv:1302.4389*, 2013.

[65] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[66] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[67] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.

[68] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[69] G. Robin, "Study of genetic algorithms behavior for high epitasis and high dimensional deceptive functions," in *Evolutionary Computation*. InTech, 2009.

Chapter 4

A Comparison of Sexual Selection versus Random Selection With Respect to Extinction and Speciation Rates

4.1 Introduction

Within biological literature, a distinction exists between sexual selection and random mating (panmixia), sometimes known as the null model. Strictly speaking, random mating as a null model assumes an infinite number of mates for a female. Realistically, however, there is simulation evidence that random finite mating is possible [1]. Moreover, ample evidence from empirical studies supports the occurrence of random mating in a number of species. Authors in [2] found panmixia to occur in the Euopean eel (Anguilla Anguilla L.) and proposed it to be likely within other marine species. Using genetic evidence of patterns of differentiation in a migratory species

of waterbird (Pelecanus Erythrorhynchos), the researches in [3] were able to confirm random mating in this species. Further, [4], using genetic evidence (12 polymorphic microsatellite markers), found panmixia to be likely within a species of halibut (reinhardtius hippoglossoides). Authors in [5] found evidence of panmixia in a deep sea fish, Antimora rostrate.

As observed by [6], sexual selection is a relatively poorly understood concept. They define sexual selection as intra-specific reproductive competition [6]. Further, the authors proposed a distinction between natural selection and sexual selection, where the latter is variance due to mating success and the former involves variance with respect to other aspects of fitness [6]. An important issue with respect to sexual selection is the so-called good genes hypothesis, which is the idea that females choose males with good genes leading to fit offspring. This hypothesis is based upon the assumption that males manifest evidence of genes though secondary phenotypic traits such as coloration [7]. [7] performed a meta-analysis and discovered a marginally significant correlation between offspring survival and male secondary traits that attract females during mating, thereby indicating that good genes play at least a small role in sexual selection. [8] studied Pronghorn (Antilopa Americana) females as they engage in selective mating. While they concluded that secondary male traits may play a minor role in selection for good genes [8], they also concurrently admitted that the good genes hypothesis remains an open question.

In our study, using individual-based computer simulations, we compare sexual selection and random mating with respect to speciation rate and extinction rate. In cases of sexual selection involving female selection of males with good genes, thereby possibly conferring a fitness advantage on offspring, it may be reasonable to anticipate the extinction rate of species with sexual selection to be lower than for species that mate randomly. However, the question concerning the relative fitness advantage and

extinction rates of either random mating or sexual selection remains unresolved in the literature [9]. It is our hope that our computer simulations will shed some light on this issue by determining the conditions under which random mating and sexual selection contribute to the extinction rate. It is not being claimed that there is any unresolved inconsistency between random mating and sexual selection in terms of extinction rate.

Using numerical simulations, the authors of [9] argued that the concept of sexual selection resulting in lower or higher extinction rates (compared to random mating) is dependent upon the female responsible for the survival of the species gaining benefits from selective mating without suffering costs. The authors argued that scenarios which require the female to bear the costs of selective mating without gaining the benefits, would lead to higher extinction rates versus random mating scenarios. However, within scenarios in which the female gains the benefits of sexual selection while the male bears the costs, the extinction rate for the species would be lower than within random mating [9]. [10] provided empirical evidence demonstrating sexual selection to actually counteract extinctions of bulb mite populations Rhizoglyphus robini. They created 100 small populations of mites, each with 5 males and 5 females, where 50 populations required monogamy and 50 populations allowed sexual selection [10]. They discovered monogamous populations to display a significantly higher extinction rate, as opposed to the sexual selection populations [10]. Along the same lines, [11] found lineages of the flour beetle (Tribolium castaneum) derived from populations with strong sexual selection to display fitness despite inbreeding, thus suggesting that sexual selection serves as protection against extinction. Lineages derived from populations with weak sexual selection or no sexual selection, however, experienced lower fitness and became extinct after 10 generations [11]. Again, it is important to emphasize here that we are not pitting sexual selection against random mating. Rather, we are citing evidence for differences between them in terms of extinction rates.

Conversely, a study by [12] found that sexual selection in 1030 bird species increases the risk of extinction while possibly promoting speciation. Moreover, [13] found no evidence of a relationship between extinction rate and sexual selection after examining data of body masses of 1007 species of mammals.

A recent individual based modeling study predicted sexual selection to be capable of potentially increasing the probability of extinction and, in other cases, decreasing the probability of extinction [14]. According to the authors, sexual selection decreases the probability of extinction in larger populations while increasing the probability of extinction in small populations [14]. The aim of our current simulation study is to shed further light on this issue.

A related issue examines the presence of a relationship between sexual selection and speciation rate. Similar to the debate regarding the possible connection between sexual selection and extinction rate, there exists no clear consensus in the literature regarding the connection between sexual selection and speciation rate. Supporting evidence has been presented by a number of studies that suggest sexual selection as a driver of speciation. For example, [15] studied 84 speciation events in 23 species of passerine birds. They concluded that sexual selection, combined with male-male competition, hastened the evolution of pre-mating reproductive isolation, thereby driving speciation [15]. Further, [16] found animals that participate in bioluminescent courtship displays (suggesting sexual selection) to display a high rate of speciation. Additionally, the author of [17] found sexual selection generated by sensory drive to contribute to reproductive isolation and, hence, speciation in threespine sticklebacks (Gasterosteus spp.). Refuting evidence has also been published. [18] conducted a meta-analysis to determine the effect size of any possible correlation between sexual selection and speciation rate. The authors found a small, albeit significant, positive correlation between sexual selection and speciation rates [18]. [19] argued that there

exists no supporting evidence that sexual selection drives speciation in birds. They employed data for 1030 bird species across 467 genera to compare speciation rates between species displaying strong sexual selection with species displaying random mating [19]. They found that although species richness varied across the genera, there was no correlation between species richness and sexual selection [19].

In an individual based modeling study, [20] found sexual selection to reduce reproduction in immigrants, thereby reducing speciation by hybridization. Authors in this study assumed hybridization to be a contributing mechanism for speciation, and found intraspecific sexual preferences to appear to reduce species divergence. Further, [20] suggested that sexual selection is not sufficient to independently promote speciation. In contrast, computer simulations employed by [21] demonstrated that genetic drift can work in conjunction with sexual selection to promote speciation. Moreover, [22] , using individual based computer simulations, found that sexual selection acts in concert with natural selection to promote reproductive isolation and eventual speciation. However, [23], using agent based modeling computer simulations of sexual selection, found sexual selection to independently initiate speciation and contribute to population diversity. Thus, even in terms of the simulation literature, there is a measure of disagreement regarding the role of sexual selection in speciation. Using our individual based modeling computer simulations, we hope to provide additional understanding of the possible connection between speciation rate and sexual selection.

Species turnover and species diversification rates are closely related to the concepts of extinction and speciation, respectively. Turnover rates can be defined as the interplay between species extinction and the advent of new species through speciation or immigration, with speciation tempering extinction [24] [25] [26]. Authors in [26] report that dichromatic species of birds, with higher rates of sexual selection due to color displays, have a 23% higher local extinction rate and a 25% higher turnover rate

than monochromatic (only black and white vision) bird species. This is consistent with the findings of [12]. Diversification rate can be defined as the difference between speciation rate and extinction rate as outlined in Magallon and Sanderson (2001) [27] and in [28]. Similar to [15], [16] and [17] examined the relationship between sexual selection and speciation rate, Kazancioglu et al. (2009) [29] reported that sexual selection (as evidenced by dichromatism) accelerates diversification rates in parrotfishes (Scaridae). Moreover, [30] collected empirical data from a wide range of animal taxa employing sexual selection. The authors found that taxonomic families with strong sexual selection on males demonstrated higher levels of species richness, as opposed to taxa without strong sexual selection on males [30]. On the other hand, [31] found no significant difference in diversification rates between dichromatic species of cichlid fish and monochromatic species of cichlid fish in Lake Tanganyika.

Further, in an individual based modeling simulation study employing a genetic algorithm, [32] reported that sexual selection initiates speciation, along with biodiversity, by creating new fitness peaks. Additionally, a combined phylogenetic and computer simulation study demonstrated sexual selection to increase genetic diversity within populations of jumping spiders (Habronattus pugillis), thereby suggesting a link with speciation and species diversity [33].

There are several empirical studies that consider the possible mechanisms relating sexual selection with speciation and extinction rates. For example, [34] studied X. birchmanni-X. malinche hybrid zones in river systems in Mexico. The authors found that assortative mating (a form of sexual selection) leads to higher rates of speciation by inducing reproductive isolation, which is a key mechanism in speciation [34]. [35] found that although in certain species of Nematodes, unisexual reproduction (and hence the absence of assortative mating) led to reproductive isolation and higher species diversification, in the plant species Mimulus self-fertilization led to lower levels

of reproductive isolation and hence species diversification. In a recent review article, [36] argue that although assortative mating can in some cases lead to higher resistance to pathogens (and hence lower extinction rates), in other cases, assortative mating reduces genetic diversity thereby lowering resistance to pathogens.

In this study, using individual-based computer simulations, we will test the following hypotheses regarding sexual selection and speciation and extinction rates.

H1: There is a significantly higher rate of speciation in populations with strong sexual selection versus those with weak or no sexual selection.

H2: There is a significantly lower rate of extinction in populations with strong sexual selection versus those with weak or no sexual selection.

It is important to note that in testing these hypotheses, we are not myopically ruling out a variety of ecological factors contributing to speciation and extinction rates such as habitat, predation, population size, body size, resource levels, and so forth. These factors admittedly play a role with respect to extinction as discussed in [37] and with respect to speciation as discussed in [38]. In fact, our rules performed extraction using decision trees after the runs to appeal to ecological factors, such as body size, in predicting extinction and speciation rates (see below). Moreover, the initial time steps of all runs involved similar parameter values for ecological factors such as predation, habitat, grass availability, and more in order to isolate the independent variables of sexual selection and random mating.

Further we will test the following hypotheses regarding sexual selection, species turnover, and species diversification:

H3: There is a significantly higher species diversification rate in populations with strong sexual selection versus those with weak or no sexual selection.

H4: There is a significantly lower turnover rate in populations with strong sexual selection versus those with weak or no sexual selection.

With respect to H3, the prediction of higher rates of extinction in populations with random mating can involve offset by higher rates of speciation. High rates of speciation, as a counter to high extinction rates, were discussed by the author of [39]. He observed that high speciation rates were the only way for Neogene pectinaceans to avoid extinction. On the other hand, using individual based modeling, [40] showed that a low speciation rate may actually confer resistance to extinction. Our regression analysis will help resolve this issues. Finally, we will use machine learning (employing decision trees) to extract rules predicting extinction rates and speciation rates both for sexual selection and random mating. These rules will help to discern whether female selective patterns, or lack thereof, predict extinction rates and speciation rates.

4.2 Materials and Methods

4.2.1 EcoSim

EcoSim [41] is an artificial ecosystem comprised of a three level food chain. Predators occupy the third trophic level, followed by prey at the second trophic level, and grass as primary producers. As an individual based modeling system, EcoSim is designed with a bottom-up approach, whereby the system defines evolvable individual (prey and predator) behaviors and exhibits emergent properties of a natural ecosystem at a macroscopic level *. The source code of EcoSim can be downloaded from github.com/EcoSimIBM.

The virtual world of EcoSim consists of 1000x1000 cells represented as a torus, where each cell in the world has the potential to contain an unlimited number of prey and predators, and limited amount of grass. Every individual has a unique behavioral model that is coded in its genome and is therefore subject to evolution. In EcoSim,

*The source code of EcoSim can be downloaded from github.com/EcoSimIBM

predators can hunt prey, which results in an allotment of meat in the cell proportional to the size of the prey killed. The location of individuals and amount of resources dynamically vary in time as a consequence of the interactions between individuals and the diffusion mechanism for grass [42]. EcoSim has been used to investigate a wide range of ecological properties such as speciation rate [43], extinction of species [44], evolution of prey individuals under the stress of predator pressure [45], and ecological risk assessment of Polychlorinated biphenyls [46].

The genome of an individual consists of two sections; the physical genome represents physical features and the behavioral genome codes for the behavioral model of an individual. Hence, every individual has its own behavioral genome that guides its behavior. The behavior of each individual is implemented using a Fuzzy cognitive map (FCM) [47]. The FCM of an individual is a directed graph where nodes represent concepts and edges between nodes represent the relationship between the concepts. Concepts can be of three types: sensitive concepts (such as perception of distance from predator, perception of self-energy levels, etc.), internal concepts (such as fear, satisfaction, etc.), and motor (or action) concepts (such as reproduce, escape from predators, etc.). The edges between two concepts have weights denoting the influence of one concept upon another. A positive weight models an excitatory effect from one concept to another, and a negative weight models for an inhibitory effect. Activation of sensitive concepts influences activation of internal concepts, and activation of internal concepts influences activation of motor concepts . Specific positive and negative feedback loops can also evolve.

One possible action for prey and predators is reproduction. Reproduction requires two individuals and is successful if these two individuals are located in the same cell, have enough energy, and are genetically similar. The physical genome distance between two individuals is calculated as sum of differences between values of their

physical genome attributes, after the difference is normalized based on the range of the attribute. The physical genome distance is added to the evolutionary distance of two individuals, which is calculated based on differences between the values of the behavioral genome (coding for edge weights within the FCM). This results in a total measure of dissimilarity between two individuals. If this dissimilarity is less than a certain predefined threshold, the individuals are able to mate. For all experiments in this paper, this threshold is set as 16 for both prey and predator.

A high degree of genetic dissimilarity causes reproduction to fail. Conversely, a unique new individual is born as a result of successful reproduction. The resulting offspring are assigned a distinctive genome inherited form both the parents with possible mutations. A group of individuals with similar genetic characteristics is assumed to be of the same species. A species is associated with a representative genome equal to the average genome of its members. A speciation threshold is used to determine the occurrence of speciation. When the difference between the genome of two individuals of the same species is greater than the speciation threshold, a speciation occurs. Thus, when populations genetically diverge due to evolution, speciation events can be generated. A 2 mean clustering algorithm is applied to the genome of the individuals of the species to split it in two sister species [48].

Individual Attributes

Individuals in EcoSim, whether prey or predator, possess physical attributes. Prey with faster running ability are better equipped to escape from predators, while prey with a slower running ability are hunted. Additionally, individuals that can sustain a steady energy level are expected to live longer. The attributes or physical traits in question are of two types in EcoSim: those that are inherited by individuals at birth (maximum energy, maximum age, vision range, maximum speed, minimum

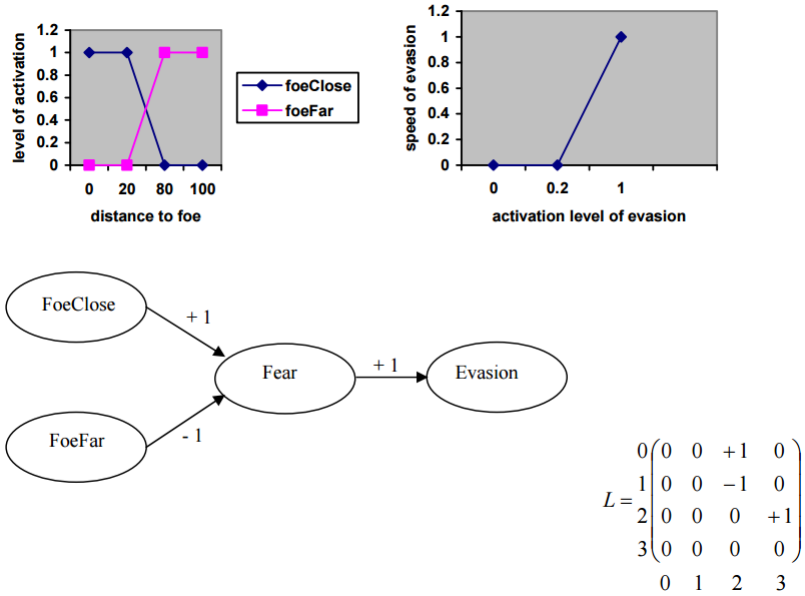


Figure 4.1: An example of a simple FCM in which activation of foeClose (proximity to predator) and foeFar (distant predator) is given by fuzzification of these concepts, depending on the distance of prey from predator. In the fuzzification process the real value of the sensory concept (say predator is 5 cells away) is converted to a fuzzy value (say a decimal number from 0 to 1). The speed at which prey evade is given by defuzzificaton of evasion concept, where the reverse of fuzzification happens - the fuzzy value is converted to a real scalar value. The L matrix is an nxn matrix showing influence of one concept on another; where 0 denotes foeClose, 1 denotes foeFar, 2 represents fear and 3 represents evasion. Activation levels of motor concepts in EcoSim dictate what action an individual will take next and the defuzzification of the activation level provide the intensity of the action. For example if evasion concept is activated, the defuzzification of evasion concept gives the speed of evasion. [41]

reproduction age, State of birth, defense, and cooperative defense) and those that are acquired by an individual during its lifetime (such as energy at a given time step, age, speed, and strength).

Operational Definitions of Turnover Rate and Diversification Rate

Following the lead of [26], in equation 4.1, turnover rate (TR) is defined as the extinction rate (ER) divided by the speciation rate (SR):

$$TR = ER/SR \tag{4.1}$$

Further, following the lead of [27], in equation 4.2, diversification rate (DR) is defined as the difference between speciation rate (SR) and extinction rate (ER):

$$DR = SR - ER \tag{4.2}$$

Energy and Strength

The world in EcoSim contains grass as the primary producer, which is initially uniformly distributed. Grass can then grow up to a user defined limit, diffuse from one cell to another, and is finally consumed by prey to obtain energy for sustenance. Predators hunt prey for meat or scavenge dead prey for energy. The maximum energy of an individual (energyMax) is part of its physical genome and is constant for an individual. Cumulative energy obtained by consuming grass or meat is capped at the energyMax of an individual. The energy obtained by individuals contributes to an individual's strength. The strength of an individual is dependent on an individual's age, current energy, and maximum energy. For prey individuals, if the energy of an individual is greater than or equal to 1/3 of its maximum energy and if this individual is young or old, its strength is equal to 2/3 of maximum energy. If the energy of an

individual is greater than or equal to 1/3 of maximum energy and if the individual is neither young nor old, its strength is equal to maximum energy. However, if the current energy of an individual is less than 1/3 of maximum energy, then its strength is equal to the sum of 3/2 of current energy and 1/2 of maximum energy. An individual is considered young if its age is less than the reproductive age (repAge), and an individual is considered old if its age is greater than maximum age of the individual minus the reproductive age (maxAge - repAge). The average repAge for prey is 6 time steps, while the maxAge for prey is 46 time steps; these values change across generations of individuals as a result of evolutionary processes modeled within EcoSim.

The energy obtained by individuals is spent at each time step to carry out various actions necessary for survival and procreation. Equation 4.3 gives the energy spent by prey at each time step. Table 4.1 gives initial values for some of the parameters.

$$\begin{aligned} Energy_{prey} = 1.2 \times ArcPenalty + \frac{strength \times (speed)^2}{4000} + \left(\frac{MaxEnergy}{5}\right)^{0.75} \\ +(Vision \times 2.5)^{0.75} + (defense \times 5)^{0.75} + (Max(0, 8 - repAge))^{2.3} \end{aligned} \tag{4.3}$$

$$ArcPenalty = max((nbArcs - threshold)^{0.75}, 1)) \tag{4.4}$$

In equation 4.4 arcPenalty accounts for energy consumed by the FCM of an individual based on the individual's FCM complexity. NbArcs are the number of edges within the FCM of an individual (which includes connections between sensory, internal, and motor concept nodes). In other words, an individual is imposed a higher energy cost for having a complex behavioral model (FCM) within EcoSim, as compared to individuals with a simpler behavioral model (lesser edges within FCM). The

cost is imposed only when the number of edges/arcs go beyond a predefined threshold.

In equation 4.3 vision refers to the extent (number of cells) for which individuals can see. Prey individuals can defend themselves from predators, and the value of this faculty is given by the defense parameter. RepAge is the minimum age at which individuals can begin to reproduce. This paper presents results for prey individuals only, thus only the energy function of prey has been provided. The bioenergetics of prey individuals outlined above is consistent with the modeling used in a seminal paper on this subject, [49], for agent based models. In [49], the energy used by prey is a function of a number of variables such as food acquisition, reproduction, defense against predators, and body mass (represented by MaxEnergy) similar to our modeling. Equations 4.3 and 4.4 account for the (non-reproductive) outflow of the energy of an individual at a given time step and is subtracted from the current energy of the individual.

Reproduction

By default, females mate with randomly chosen males located in the same cell as the female. For a mating action to progress, several conditions must be established: the male must not have acted in that time step, the male must also be willing to reproduce, energy of the male must be greater than 12.5 percent (arbitrarily chosen) of its maximum possible energy, and the male must be in the same cell as the female. Additionally, females are only able to mate with males within a certain genetic mating distance. If any of the conditions fail, the female proceeds to find another male for mating, in the same time step, until an attempt to reproduce successfully with every male in the cell has been made. A penalty of 1 percent of maximum energy is paid by the females for every failed mating attempt, irrespective of the reason for failure. It is also possible that the female is unable to find any suitable mate within a given time

step and, in that case, the reproduction action becomes a futile attempt incurring loss of energy as the only effect.

In good genes mate selection experiments, females attempt to mate with males, in the same cell, by descending order of male strength. The premise relies on the assumption that greater strength is an indicator of better physical constitution and abilities. However, finding a strong male would require the female to spend more energy in search of mate. Therefore, for every failed reproduction attempt, females spend 2 percent of their maximum energy. Moreover, the strong male is selected before the female finds out whether other conditions are favorable; thus the probability of a successful reproduction, given that the male is strongest and most able to reproduce (all other conditions are favorable), is reduced.

When a reproduction action is successful, an offspring is born. The sex of the newborn is selected randomly with equal probability of being male or female. The genomes of the parents are crossed over and mutated to form the genome of the offspring. The offspring inherit the perception layer within the FCM from the parent of same sex as the offspring. The rest of the edges in the FCM of the newborn are randomly picked from either parent. In order to model simple linkage, alleles corresponding to the edge values of the FCM are transferred by blocks; for a given node in the FCM, the values for each of the edges connected to the node are transferred together from one parent to the offspring. For each such block of FCM edges, the parent from which the block is taken is chosen randomly. Hence, for a given node, there is no recombination among edges; recombination occurs at the level of edges of blocks.

The physical genome of the offspring is also a random combination from the physical genome of the parents. This is true for the whole physical genomes, with exception of maximum energy (ME) as shown in equations 4.5 and 4.6. Refer to table 4.1 for

initial values of ME.

$$ME_{femalenewborn} = \frac{(5 \times ME_f + ME_m)}{6} \tag{4.5}$$

$$ME_{malenewborn} = \frac{(5 \times ME_m + ME_f)}{6} \tag{4.6}$$

In equations 4.5 and 4.6, m and f refer to male parent and female parent respectively.

The initial energy of an offspring at birth is governed by the State of Birth (SOB) (table 4.1 gives initial values of SOB) of the parents - which is a fraction of energy contributed by the parents to the offspring at the time of reproduction.

$$Energy_o = \frac{ME_f \times SOB_f + ME_m \times SOB_m}{100} \tag{4.7}$$

Energy of both the parents is reduced following a successful reproduction. The new energy of the male parent is defined by equation 4.8.

$$NewEnergy_{maleparent} = \frac{SOB_m}{SOB_m + SOB_f} \times Energy_o \times 1.05 \tag{4.8}$$

If the offspring produced is a result of the first pregnancy of the female parent, the female parent incurs an additional First Pregnancy Penalty (FPP) where Energy_f is the energy of the female before reproduction, shown in equation 4.9. There is biological justification for this bioenergetics model of reproduction that imposes a penalty for primiparity (first birth). In [50], it was reported that first pregnancies involve more complications than second pregnancies in the case of humans (homo sapiens sapiens), such as hemorrhaging and perinatal death. Further, it Was reported in [51] that guinea-pigs' (Caviaporcellus) transfer of maternal energy to offspring during the first

Table 4.1: Initial values of the parameters of submodels related to prey individuals

Parameter	**Prey Female**	**Prey Male**
State of Birth (SOB)	16	12
Max Energy (ME)	2000	2000
Vision	13	13
Age of Reproduction (repAge)	6	6
Maximum Age	46	46

pregnancy is less efficient than for multiparous mothers. In addition, births have been shown to be less successful for primiparous mothers than for multiparous mothers.

$$NewEnergy_{femaleparent} = \left(\frac{SOB_f}{SOB_m + SOB_f} \times Energy_o \times 1.05\right) - (FPP \times Energy_f) \tag{4.9}$$

If the offspring does not result from the first pregnancy of the female parent, the energy after reproduction is given by equation 4.10.

$$NewEnergy_{femaleparent} = \left(\frac{SOB_f}{SOB_m + SOB_f} \times Energy_o \times 1.05\right) \tag{4.10}$$

4.2.2 Machine Learning

According to [52], machine learning can be described as a field of research that deals with the automated analysis of complex data with aim to predict certain dependent variables of the data set given independent variables. The ultimate objective is to group together related data into meaningful clusters or discover patterns in a data set. Machine learning models are considered to perform well if they can generalize to new data, after the model has been generated using known data. Accordingly, the data processed by machine learning algorithms can be divided into two groups [52].

One is used for training or building the model, and the other (previously unseen data) is used to validate the model. Once a model is available that can predict unseen data with reasonable accuracy, one may also infer meaningful rules from the model with the aim of discovering the factors that influence the predicted outcome [44].

Decision Trees

According to [53], a decision tree is a tool used to describe underlying data with a set of rules that recursively divides the data into a hierarchy of such rules in a sequential manner. Thus, decision trees aid in exploring data by providing a description of significant properties of the data. Each path in a respective tree corresponds to a rule learned from the data that can be used to predict the values of dependent variables from the observed values of independent variables. These rules can then be used to discover and understand specific causal relationships between these variables.

Given the scope of the current paper, the discussion is restricted to univariate classification trees with ordered decision variables.

In most univariate decision tree algorithms, partitioning of the data is done by splits based on a single variable of the form $x \leq c$ (where x is a non-categorical variable and c is a constant). This implies that if x is less than a certain value, data is assigned within one group. When x is greater than a certain value, the data is allocated to a different group. This point of decision is referred to as a (decision) node and is a point where the tree either branches into more nodes or ends in a leaf (containing a class). These nodes can be easily converted into IF-THEN rules, which enhance the interpretation of a decision tree model.

Generally, at the level of a node, an exhaustive search is employed to find the variable x and the constant c; with an objective to optimize some measure of node impurity such as *entropy* I_m [54], given by equation 4.11, for classification trees.

Decision trees are built top-down starting from a root node, and then partitioning the data into sub-spaces with as much homogeneity as possible. Namely, after a split, each child node contains instances of similar class. The entropy measure in equation 4.11 gives the homogeneity of a node. If all instances in the node are of same class, the equation evaluates to 0. Otherwise, when the instances are equally divided by two classes, the entropy value is 1.

For growing or induction of trees, greedy search techniques exist that employ heuristics to find smaller trees in less time.

$$I_m = -\sum_{i=1}^{k} p_m^i log_2 p_m^i \tag{4.11}$$

The research presented here uses a special decision tree algorithm known as C4.5 [55]. C4.5 uses gain ratio, which is an entropy based measurement, to test purity of a node. Authors in WEKA's [56] use J48 algorithm, an implementation of C4.5 in Java programming language to model and validate decision trees.

Variational Autoencoder EDA as a Wrapper Method for Feature Selection

Machine learning algorithms work to extrapolate knowledge from massive datasets. The reliability and interpretability of the knowledge gained by machine learning analysis of the datasets, however, can be affected by presence of irrelevant, redundant and superfluous information (features) present in such datasets [57] . Feature selection is a class of machine learning algorithms that is used to detect and expunge the dataset of such extraneous features before the predictive model is built. For example, if a dataset contains samples described by 10 features and only 3 of these features are useful in classifying the dataset into one category or another, then feature selection algorithms remove the other 7 features so that a more concise and efficient classifier can be built using these 3 features only.

Wrapper methods are a type of feature selection algorithms where the model building step (such as generating a decision tree) is integrated within the feature selection step [58]. This differs from the filter method in which the feature selection is performed as a preprocessing step before the model is built. Wrapper methods define a combinatorial optimization search procedure in the complete feature space that return subsets of features to be evaluated based on the predictive model. In the evaluation phase, these subsets of features are used to train and test a specific classification model (consider a C4.5 decision tree, for example), and the accuracy of the decision model generated from this subset is used to judge the worthiness of the feature subset. Thus, to search the feature space for the best subset, the search is wrapped around the classification model.

Estimation of distribution algorithms (EDA) [59–62] are metaheuristics that aid in combinatorial optimization. EDA maintains a population of promising candidate solutions. In a selection step, the candidate solutions with the best fitness are selected from the population, and thereafter a probabilistic generative model is built from these promising candidates to generate the next generation of promising candidates. Fitness of the candidate solutions are defined by problem specific criteria that evaluates the candidate as a whole, based on the values of each constituent feature in the candidate solution. In our case, the fitness function is the predictive models accuracy. This process repeats itself until the optimal solution is obtained or the current best solution has a fitness which is deemed as sufficient. The exact combination of assignments to variables producing a higher fitness is complex to determine. EDAs have been used successfully as a wrapper method for classification tasks in [63] and [64]. In [64] Constrained Model Search Space Bayesian Optimization Algorithm (CMSS-BOA) [65] was used as a wrapper method with C4.5 Decision trees. When EDAs are used as the Wrapper Method, the fitness is evaluated based on the predictive accuracy of the

classifier.

Variational Autoencoder EDA with Population Queue (VAE-EDA-Q) [66] which uses variational autoencoders [67] [68] to model the population. This method has been found effective at exploring latent continuous search spaces. VAE-EDA-Q is able to explore continuous latent space representation of the features, which allows randomized sampling and smooth interpolation over the manifold of the high-quality solutions in the feature space.

4.3 Simulations

Ten independent runs of EcoSim *, were performed on SHARCNET (www.sharcnet.ca) for each of the following two experiments - Sexual Selection strategy and Random Mating strategy. Each of the runs was allowed to pass five thousand time steps. One generation in EcoSim corresponds to minimum reproductive age (repAge) and is approximately equal to 6 time steps. The experiments were performed on AMD Opteron systems with 2.2 Ghz clock speed, with four cores, and 8 GB of RAM.

For a given run, population wide average results (containing information about speciation rate, average energy, mating distance, extinction rate, etc.) are available for prey and predator individuals respectively, irrespective of the species to which the individuals belong. In order to analyze the commonality of results between the two experiments, we averaged the results from 10 runs for each of the two experiments instead of focusing on isolated results from each run, which yielded two sets of results: one for sexual selection (average of ten runs) and one for random mating (average of ten runs). The world and individuals of EcoSim take nearly one thousand time steps

*Source code of EcoSim is available from github.com/EcoSimIBM for readers interested in replicating the experiments.

to stabilize into a sustainable world containing genetically fit individuals. Therefore, the first one thousand time steps were removed from the two result files.

Extinction rate in EcoSim is measured as the number of species that became extinct in a given time step divided by the total number of species at that time step. Similarly, speciation rate is the number of new species created in a given time step divided by the total number of species at that time step.

The one way ANOVA test described by Kruskal Wallis [69] on ranks was performed between the speciation rate for sexual selection and the speciation rate for random mating, as well as between the extinction rate for sexual selection and the extinction rate for random mating with an α value (or significance level) of 0.05. The Kruskal Wallis test was also performed on species turnover rates and diversification rates. In either case, the null hypotheses was assumed to be that there was no significant difference in speciation rate or extinction rate between the two groups of sexual selection (good genes mating) and random mating.

As mentioned previously, the machine learning experiments were aimed at predicting the variation in speciation rate and extinction rate for a given mating strategy. Hence, the speciation rate and extinction rate needed to be transformed from non-categorical to categorical variables in order to change the regression problem (predicting values) to a classification problem (predicting class labels). The interest was to predict, at a population wide level, the aspects of high and low speciation rate, and high and low extinction rate.

To transform the speciation and extinction rates, the data was divided in to three parts based on values of speciation rate (for classifying speciation rate) and values of extinction rate (for classifying extinction rate) such that the number of instances in each of the three parts was equal. The objective was to divide the data set into three classes (HIGH, MED, and LOW) based on either speciation rate discretized

Table 4.2: Mapping speciation rates and extinction rates to HIGH-LOW classes

Class	**Sexual Selection**	**Random Mating**
Speciation HIGH	Speciation rate > 0.00787	Speciation rate > 0.01010
Speciation LOW	Speciation rate < 0.00513	Speciation rate < 0.00546
Extinction HIGH	Extinction rate > 0.00790	Extinction rate > 0.00933
Extinction LOW	Extinction rate < 0.00512	Extinction rate < 0.00525

to speciation level or extinction rate discretized to extinction level. The analysis of causes for a significant increase or decrease in speciation and extinction rates was of interest; our focus, therefore, centered on predicting the instances in the partitions where speciation/extinction was labeled as "LOW" and in the partition where speciation/extinction was labeled as "HIGH". Total instances in the partition labeled “MED” were discarded, as they were not useful for our investigation.

Table 4.2 shows the thresholds for which a given speciation or extinction rate is mapped to a HIGH-LOW class. The values that fall between these thresholds are classified as medium or MED. The results presented in this paper are based on approximately 3900 instances, or data points, for sexual selection and 3900 instances for random mating (after discarding the MED values from result sets). Each instance, or data point, corresponds to a time step in the EcoSim simulation. Consequently, each of the 3900 instances can be classified as Speciation HIGH or Speciation LOW, for Speciation level class. The same 3900 instances can also be classified as Extinction HIGH or Extinction LOW, for Extinction level class, based on the values in table 4.2.

We used C4.5 for predicting speciation and extinction levels using 10-fold cross validation, and the results presented in the next section are an average of all the folds.

Sensitivity analysis was performed to discover the extent to which the key findings potentially varied based on the penalty imposed on females for failed reproductive efforts (equation 4.12 where N is number of failed reproduction attempts, p= 0.01 for random mating and p = 0.02 for sexual selection). A secondary set of experiments

on sensitivity of the reproductive penalty parameter for sexual selection runs was performed.

$$Energy_{new} = Energy_{old} - [N \times p \times Energy_{max}] \tag{4.12}$$

In the above mentioned experiments, a coefficient parameter c was added to the penalty as given in equation 4.13 which was varied from 0.50 to 1.25, in steps of 0.25. Thus, 15 independent runs were performed for each of the five coefficients, with a run time of 4500 time steps. For a given coefficient, the average of speciation and extinction rates were calculated for 15 runs. Furthermore, the rates were averaged for 30 time steps, in steps of 50 time steps, to avoid inherent temporal correlation (see figure 4.2).

Thereafter, Kruskal Wallis Analysis of variance was performed on each of the five sets of results to establish whether any change within the threshold of coefficient of the penalty parameter would have any significant impact on speciation or extinction rates that would affect the key findings presented here.

$$Energy_{new} = Energy_{old} - [c \times N \times p \times Energy_{max}] \tag{4.13}$$

4.4 Results and Discussion

The graph shown in figure 4.3 shows the similarity of extinction rates when the coefficient of penalty imposed on female prey for reproductive failure is varied from 0.5 to 1.5 in steps of 0.25.

Kruskal Wallis analysis revealed no significant difference between the extinction rates for various treatments (p value of 0.492, for $\alpha = 0.05$) for female prey in sexual

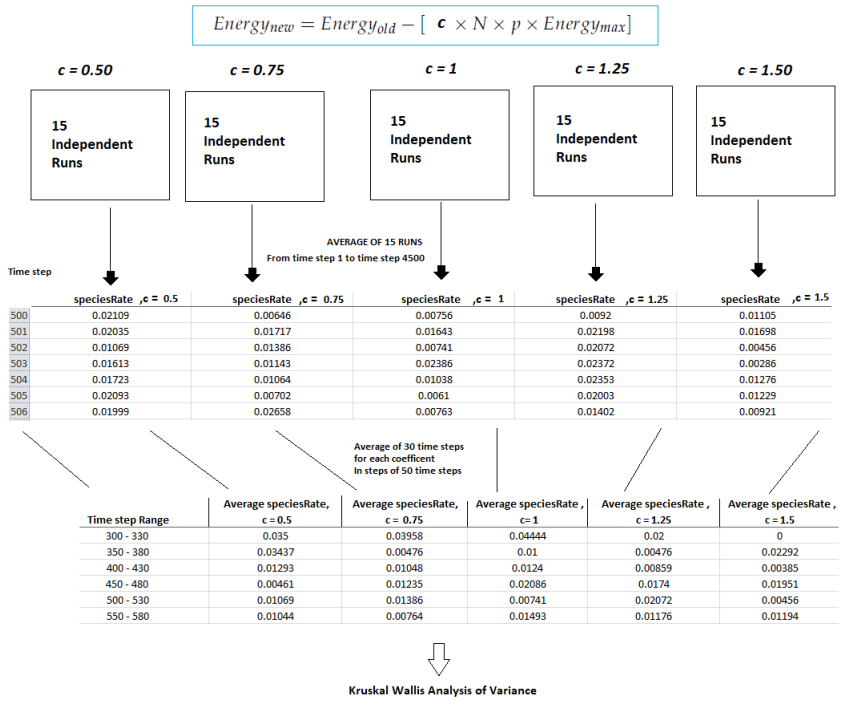


Figure 4.2: Measuring whether statiscally significant differences exist in species rates, for varying values of reproductive penalty parameter based on coefficient c. Sensitivity anal- ysis experiments for extinction rate has same setup.

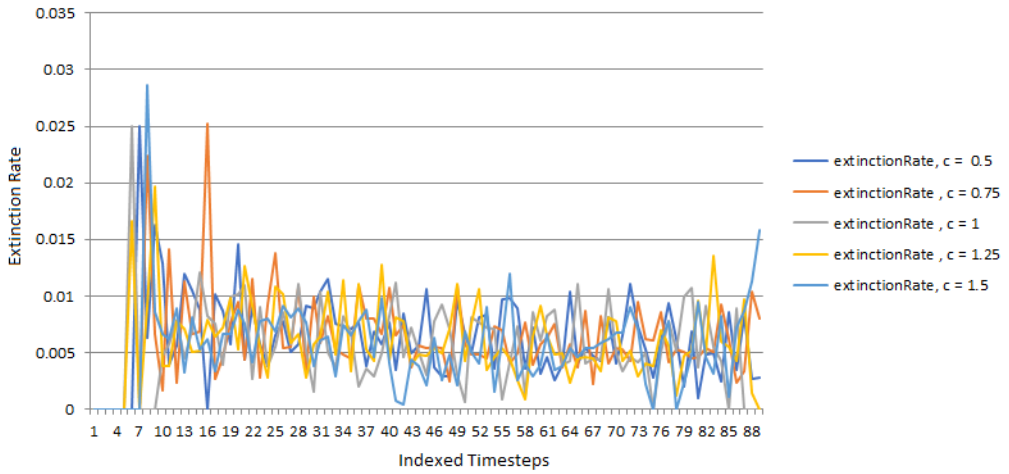


Figure 4.3: Extinction rates when the coefficient of penalty imposed on female prey for reproductive failure is varied from 0.5 to 1.5 in steps of 0.25. The differences observed are not statistically significant.

selection experiments (figure 4.3). Similar results were obtained for speciation rates; the graph in figure 4.4 shows the change in speciation rates when c is varied from 0.5 to 1.5. Kruskal Wallis analysis also showed no significant difference in speciation rates for different c-values (p value of 0.179, for $\alpha = 0.05$).

Hence, no statistically significant differences were observed in speciation and extinction rates when the reproductive penalty parameter was altered by a factor varying from 0.5 to 1.5. This establishes robustness of the results and discussions that follow, which are to include turnover and diversification rates.

4.4.1 Lower speciation rate in sexual selection

Recall that our first hypothesis, H1, describes a higher speciation rate for populations with strong sexual selection as opposed to populations with weak selection or random mating. As evidenced by our simulation study, the average speciation rate for the random mating group was significantly higher than the average speciation rate for the sexual selection mating group (Kruskal Wallis, p-value less than 0.0001, alpha

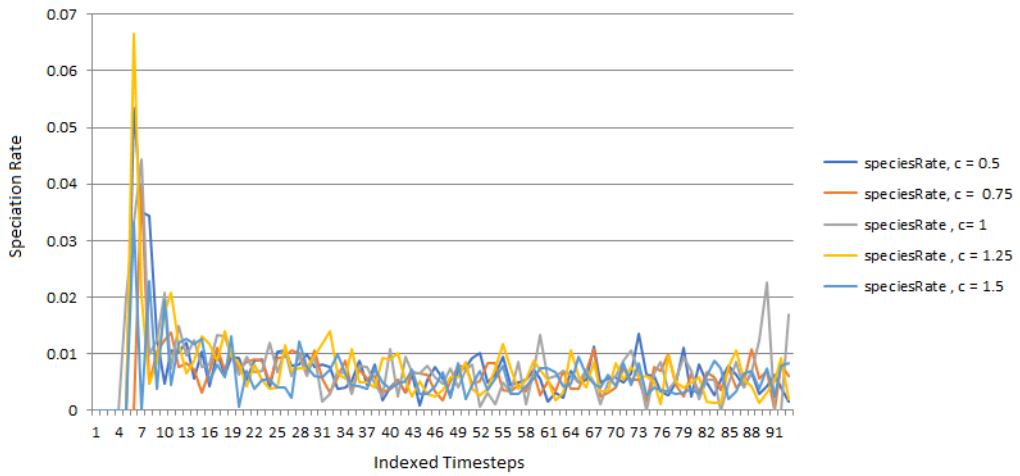


Figure 4.4: Speciation rates observed are not significant, when the coefficient of penalty imposed on female prey for reproductive failure is varied from 0.5 to 1.5 in steps of 0.25. The differences observed are not statistically significant.

= 0.05). This finding challenges H1. One possible explanation for the higher rate of speciation in the random mating group is that there is a significantly higher extinction rate for this group (see discussion in section 4.2 below) - this higher extinction rate is offset by the higher speciation rate [39].

Another possible explanation for lower speciation rates observed in populations with strong sexual selection is that sexual selection leads to lower levels of hybridization, a possibility that was considered in [20]. Moreover, this explanation is supported by an empirical study by [70] in which the authors report that sexual selection possibly impedes gene flow and, hence, hybridization between two subspecies of mice (Mus musculus musculus and Mus musculus domesticus). [70] found that male musculus mice gave preference to signals from female musculus mice and discriminated against signals from female domesticus mice, which in turn lowered hybridization. Similar findings regarding a small fish, Poecilia Mexicana, are reported in [71]. Impeding hybridization due to sexual selection would presumably impede speciation, assuming that hybridization contributes to speciation, as [72] contend.

4.4.2 Lower extinction rate in sexual selection

Our second hypothesis, H2, is that populations with strong sexual selection have a lower extinction rate than populations with either weak selection or random mating. In our simulation study, we found that there there is indeed a significantly higher rate of extinction for random selection groups as opposed to sexual selection groups (Kruskal Wallis, p-value less than 0.0001, alpha =0.05), which corroborates H2. Further, this finding of our simulation study is validated by a number of empirical studies including [10] and [11]. Authors in [40] suggest that a lower speciation rate may actually work as a shield against extinction based on the assumption that high rates of speciation tend to be selected against.

Further, as we argued above, the higher speciation rates in random mating populations may help to offset the higher extinction rates in these populations as opposed to populations with strong sexual selection. Another possible link between lower speciation rates and lower extinction rates of populations with strong sexual selection is that sexual selection reduces hybridization (as suggested above), which shields these populations from extinction in cases where hybrids have lower fitness than their parents. As [73] argue in the case of Darwin's finches, hybrids in general are less fit than the parents (except in cases of severe climatic change such as the presence of El Nino events on the Galapagos Islands) [74].

4.4.3 Species Diversification Rates

The reader will recall that the species diversification rate can be defined as the difference between the speciation rate and extinction rate as outlined in Magallon and Sanderson (2001). Our third hypothesis, H3, considers a significantly higher species diversification rate in populations with strong sexual selection versus those with weak

or no sexual selection. H3 was not corroborated by our findings. There was not a statistically significant difference (Kruskal Wallis, p-value of 0.07 and alpha of 0.05) between the diversification rates for sexual selection species versus random mating species. Although these results are in disagreement with the findings of [29], they do agree with the findings of [31] who reported no significant difference in diversification rates between dichromatic species of cichlid fish and monochromatic species of cichlid fish in Lake Tanganyika.

The findings of our simulation study are reasonable considering that within populations of strong sexual selection, the speciation rate and extinction rates are relatively low. Whereas in populations with random mating, the speciation and extinction rates are both relatively high. Thus, it is entirely possible that the differences between speciation and extinction rates in both cases would be comparable. Consider the case of a relatively high speciation rate within a sexual selection population. In light of a relevantly low extinction rate for these populations, a significantly higher diversification rate would be expected as opposed to a random mating population. This, however, is not the case.

4.4.4 Species Turnover Rates

As noted above, the species turnover rate can be defined as the interplay between species extinction and the advent of new species either through speciation or through immigration. Generally, the species turnover rate is the rate at which one species is replaced by another species, and is measured by presence and absence data of species [75]. The fourth hypothesis considered in this study, H4, is that the species turnover rate for sexual selection populations is significantly higher that the turnover rate for random mating populations. As evidenced by our study, the species turnover rate for sexual selection species is indeed significantly higher than the turnover rate for

random mating species (Kruskal Wallis, p-value of 0.03, alpha value of 0.05). These results are validated by the empirical findings of [26], as outlined above. One possible explanation of our findings regarding turnover rate is that within sexual selection populations, there exists mating discrimination against immigrants. This results in low hybridization. Whereas in random selection populations, there exists an arguably higher level of mating and hybridization with immigrants. If there is no mating with immigrants in sexual selection populations, then the immigrant species may go extinct faster than in random selection populations. The relatively high rate of absence of immigrant species due to extinction could translate into a higher turnover rate for sexual selection populations.

4.4.5 Factors driving speciation

In order to predict speciation rates, the data set was first passed through VAE-EDA-Q AVS attribute selection to remove extraneous attributes. Thereafter, the C4.5 algorithm was used to build predictive models. The two trees (one for predicting speciation level in Sexual Selection and one for predicting the same in Random Mating) had many branches leading to the inference of multiple rules (one rule per branch). The discussion, however, was restricted to the rules that provided maximum coverage - that is, the rules that were true for the maximum number of instances. There are two numbers associated with each leaf node. The first number indicates the total number of instances reaching that leaf (rule) while the second number after the '/' indicates the number of misclassified instances.

This restriction was imposed to direct interest towards the most relevant rules/branches, which is expected to remain the same when the simulation is extended to run longer (or when more runs are included in the averages). In other words, these leaves, and the rules associated with these leaves, are the most generalized ones. The leaves of

the branches with maximum coverage have been underlined with blue color in figures 4.5 and 4.6 depicting the decision tree (C4.5 model).

Speciation affected by body size in Sexual Selection

The C4.5 algorithm was able to classify the instances (into Speciation High and Speciation Low) with 80.5% accuracy, based on 10 fold cross-validation. The prediction of speciation rates by the decision tree for sexual selection indicates that when MaxEnergy (manifested in body size) is high (above a critical threshold value), the rate of speciation is high (see figure 4.5 below). On the other hand, if MaxEnergy is lower than this threshold value, though within a certain range, then the speciation rate is low (see figure 5.2 below). In other words, this decision tree implies that body size is a predictor of speciation rate for sexual selection populations. An empirical study conducted by [76] investigated the body masses of 3,235 mammal species to determine the relationship of body size to rate of speciation. The authors found that larger sized mammalian species demonstrated a higher rate of body size evolution, as opposed to smaller size species [76]. These results may, in turn, suggest a higher rate of speciation in larger bodied mammals. By extrapolating these empirical findings to our simulations, the prediction of sexually selecting populations containing individuals with higher MaxEnergy would be expected to have higher rates of speciation than populations with lower MaxEnergy since body mass is related to MaxEnergy.

Predator pressure restricting species divergence in Random Mating

The accuracy of C4.5 algorithm, in this case, was 78.7%, based on 10 fold cross-validation. The prediction of the decision tree for speciation rates for random mating indicates that defense against predators is a good predictor of speciation rates. Note that the dominant node is MaxEnergy (related to body size) as with sexual selec-

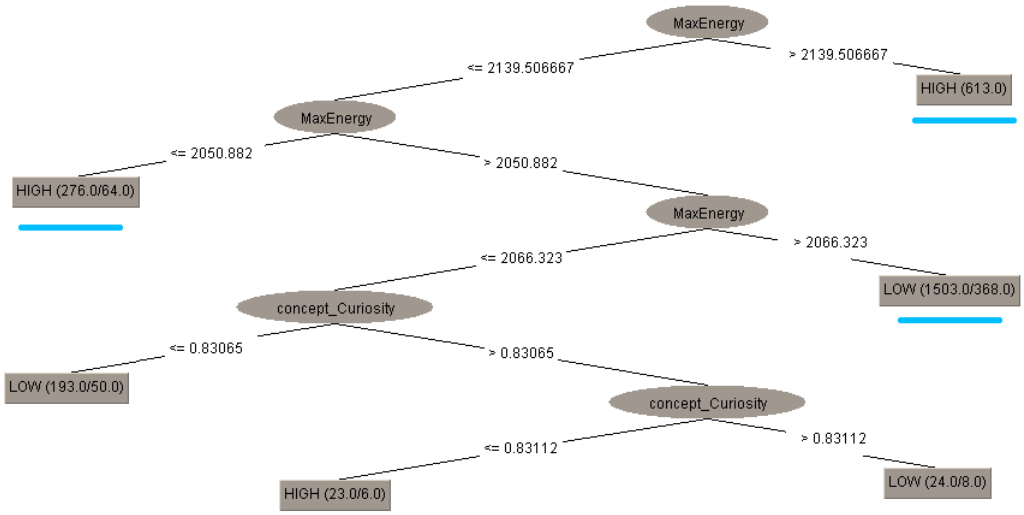


Figure 4.5: C4.5 model depicting a decision tree to predict speciation levels in Good genes mating based on threshold values of various parameters. A node represents a decision variable to predict speciation level (high or low) at the leaf. A path from various nodes to a leaf represent a conditional rule, based on the nodes in the path (decision variables). There are two numbers associated with each leaf node. The first number indicates the total number of instances reaching that leaf (rule) while the second number indicates the number of misclassified instances. The leaves of the branches with maximum number of instances have been underlined with blue color

tion. This is corroborated by an empirical study presented by authors in [76], which demonstrates another confirmation that a larger body size implies a higher rate of speciation.

For random mating, where energy expended in defense against predators is lower than a critical threshold value, the rate of speciation is high. Whereas when energy expended in defense against predators is higher than this threshold value, the rate of speciation is low (see figure 4.6 below). These results are reasonable with regard to our simulations since cases of prey expending much of their energy to evade predators are equipped with less time for reproduction. This, in turn, could slow down the speciation rate. In terms of empirical validation of these results, it was found in [77] [78] that the Western mosquito-fish (Gambusia affinis) evolves a number of traits such as smaller caudal region in lower predation settings, as opposed to members of the species in higher predation areas. This indicates species divergence between lower predation populations and higher predation populations.

4.4.6 Factors driving extinction

For predicting extinction, VAE-EDA-Q AVS attribute selection was used again before classifying the instances with C4.5 algorithm. In this case, the decision trees contained fewer nodes and, hence, the entire tree was taken into consideration in the discussions.

Higher speciation increases risk of extinction in Sexual Selection

The prediction accuracy of the C4.5 algorithm was 77.9 % based on 10 fold cross-validation. The prediction of the decision tree for extinction rate for sexual selection indicates that evolutionary distance (dist_Evol) is a good predictor of extinction rate. In particular, when evolutionary distance is above a critical threshold value, there is a high rate of extinction. Whereas when evolutionary distance is below this

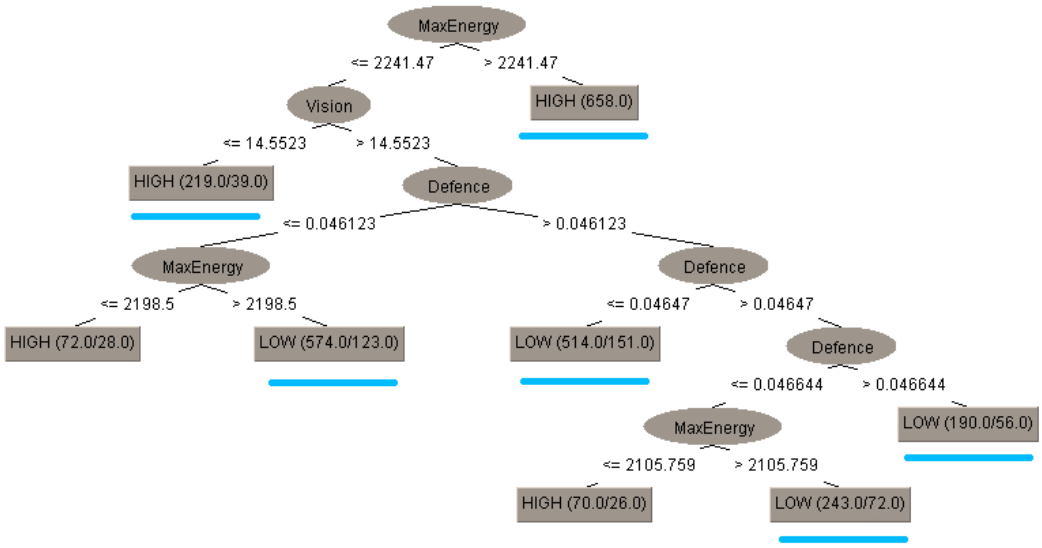


Figure 4.6: C4.5 model depicting a decision tree to predict speciation levels in random mating based on threshold values of various parameters. A node represents a decision variable to predict speciation level (high or low) at the leaf. A path from various nodes to a leaf represents a conditional rule, based on the nodes in the path (decision variables). There are two numbers associated with each leaf node. The first number indicates the total number of instances reaching that leaf (rule) while the second number indicates the number of misclassified instances. The leaves of the branches with maximum number of instances have been underlined with blue color

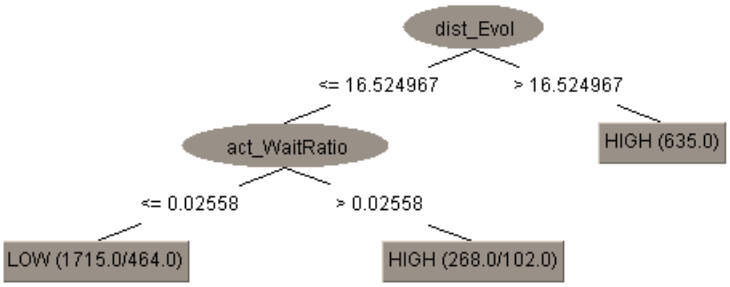


Figure 4.7: C4.5 model depicting a decision tree to predict extinction levels in Good genes mating based on threshold values of various parameters. A node represents a decision variable to predict extinction levels (high or low) at the leaf. A path from various nodes to a leaf represents a conditional rule, based on the nodes in the path (decision variables). There are two numbers associated with each leaf node.

threshold, there is a low rate of extinction (see figure 4.7 below). This concurs with our results describing species that employ sexual selection to display lower speciation rates and lower extinction rates than random mating species (which have higher rates of speciation than sexual selection species). Moreover, animal species undergoing a high rate of specialization have been described to go extinct by authors in [79]. These results suggest that high evolutionary distance resulting in speciation and attendant specialization are linked to extinction.

Larger body size increases risk of extinction in Random Mating

Finally, the prediction of decision tree for extinction rate of random mating species, which had an accuracy of 77 %, indicates that MaxEnergy (manifested in body size) is a good predictor of extinction rate. When MaxEnergy (body size) exceeds a critical threshold, there is a high rate of extinction. When MaxEnergy is below this threshold, extinction rate will be low (see figure 4.8 below). Corroborating these results is a study

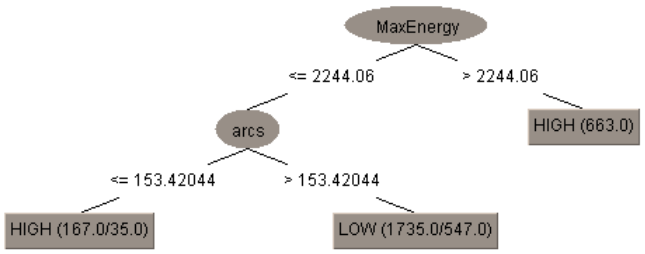


Figure 4.8: C4.5 model depicting a decision tree to predict extinction levels in random mating. A node represents a decision variable to predict extinction levels (high or low) at the leaf. A path from various nodes to a leaf represents a conditional rule, based on the nodes in the path (decision variables). There are two numbers associated with each leaf node.

of Australian mammalian species, both existing and extinct, discussed in [80]. [80] report that species with smaller body sizes are less prone to extinction than species with large body sizes. Similar results were obtained in [81] with respect to Brazilian carnivores. Extrapolating these empirical results to our simulations, leads to the prediction that random mating species with lower MaxEnergy would be expected to have lower extinction rates than random mating species with higher MaxEnergy

4.5 Conclusion

In this general simulation study, we found that species employing sexual selection have a lower extinction rate than species using random mating. This suggests that sexual selection helps to shield species from extinction, as reported in [10] and [11]. A plausible explanation of this result in our simulations is that there is lower hybridization in sexual selection species. This means that there will be a paucity of low fitness hybrids relative to random mating species. However, we found the species turnover rate for sexual selection populations to be higher than random mating populations, which may be due to the possibly faster rate of extinction of immigrants in sexual selection species habitats.

Moreover, we found a higher speciation rate displayed by random mating species than species using sexual selection. This is supported by biological theory as a higher speciation rate could help offset higher extinction rates, as discussed in [40]. Moreover, empirical evidence demonstrates preference within sexual selection species to be for mates within species classifications. This reinforces hybridization in sexual selection species to be a relatively uncommon phenomenon [70, 71]. Speciation rates within sexual selection populations, therefore, would be to be relatively low, assuming that hybridization contributes to speciation. Finally, a higher species diversification rate for species using random selection was not found. Random mating species displayed higher extinction and speciation rates, and sexual selection species displayed lower extinction and speciation rates. This implies that the differences between speciation and extinction for both groups would be comparable.

Using decision trees, we were also able to extract rules that help predict conditions under which speciation rate and extinction rate are high or low for both sexual selection and random mating. One important trait demonstrating predictive value, with respect to extinction rate for random mating, is body size (MaxEnergy). Our rules

predict that when body size exceeds a critical threshold value, there is a high rate of extinction. This is corroborated by several empirical studies [80, 81]. This is an important result since, as noted in [37], there is disagreement in the literature regarding the correlation between body size and extinction rate. Our results help to resolve this dispute, since we generated high accuracy rules that demonstrated a connection between body size and extinction rate. We also found a good predictor of extinction rate for sexual selection species to be evolutionary distance, as corroborated in Raia et al. (2016). Our findings go transcend [79] as our high accuracy rules support the link between evolutionary distance and extinction rate.

Finally, our rules show that predator defense is a predictor of speciation rate for random mating (corroborated in [77] and [78]) and that body size is a predictor of speciation rate for sexual selection (corroborated in [76]). As noted in [78], there has been no clear consensus about the role of predator defense in the driving of speciation rates. Additionally, as noted in [76], there is still debate about the role of body size in extinction rates in extant mammalian species. Once again, our high accuracy rules contribute to the resolution of both of these debates. A possible extension of this work would consider the long term evolutionary consequences of sexual selection. A similar machine learning approach to distinguish the specific evolution resulting from both sexual and random mating policies could be employed. Differing degrees of sexual selection could also be considered by changing the sexual selection policy. Long runs of our simulation could be performed to observe and compare short and long term evolutionary patterns. Our machine learning methodology could also be fruitfully applied to analyze real ecological and paleontological data, providing semantically clear rules explaining the ecological and evolutionary effects of sexual selection. However, this approach would be restricted to situation in which a large amount of data could be available.

Bibliography

[1] François Balloux and Laurent Lehmann. Random mating with a finite number of matings. *Genetics*, 165(4):2313–2315, 2003.

[2] Johan Dannewitz, Gregory E Maes, Leif Johansson, Håkan Wickström, Filip AM Volckaert, and Torbjörn Järvi. Panmixia in the European eel: a matter of time.... *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1568):1129–1137, 2005.

[3] Matthew W Reudink, Christopher J Kyle, Joseph J Nocera, Rebekah A Oomen, MCLAY GREEN, and Christopher M Somers. Panmixia on a continental scale in a widely distributed colonial waterbird. *Biological Journal of the Linnean Society*, 102(3):583–592, 2011.

[4] Denis Roy, David C Hardie, Margaret A Treble, James D Reist, and Daniel E Ruzzante. Evidence supporting panmixia in Greenland halibut (Reinhardtius hippoglossoides) in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(5):763–774, 2014.

[5] TA White, HA Fotherby, PA Stephens, and AR Hoelzel. Genetic panmixia and demographic dependence across the North Atlantic in the deep-sea fish, blue hake (Antimora rostrata). *Heredity*, 106(4):690–699, 2011.

[6] David J Hosken and Clarissa M House. Sexual selection. *Current Biology*, 21(2):R62–R65, 2011.

[7] Anders Pape Møller and Rauno V Alatalo. Good-genes effects in sexual selection. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1414):85–91, 1999.

[8] John A Byers and Lisette Waits. Good genes sexual selection in nature. *Proceedings of the National Academy of Sciences*, 103(44):16343–16345, 2006.

[9] Hanna Kokko and Robert Brooks. Sexy to die for? Sexual selection and the risk of extinction. In *Annales Zoologici Fennici*, pages 207–219. JSTOR, 2003.

[10] Magdalena Jarzebowska and Jacek Radwan. Sexual selection counteracts extinction of small populations of the bulb mites. *Evolution*, 64(5):1283–1289, 2010.

[11] Alyson J Lumley, Łukasz Michalczyk, James JN Kitson, Lewis G Spurgin, Catriona A Morrison, Joanne L Godwin, Matthew E Dickinson, Oliver Y Martin, Brent C Emerson, Tracey Chapman, et al. Sexual selection protects against extinction. *Nature*, 522(7557):470–473, 2015.

[12] Edward H Morrow and Trevor E Pitcher. Sexual selection and the risk of extinction in birds. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1526):1793–1799, 2003.

[13] Edward H Morrow and Claudia Fricke. Sexual selection and the risk of extinction in mammals. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1555):2395–2401, 2004.

[14] Carlos Martínez-Ruiz and Robert J Knell. Sexual selection can both increase and decrease extinction probability: reconciling demographic and evolutionary factors. *Journal of Animal Ecology*, 86(1):117–127, 2017.

[15] Nathalie Seddon, Carlos A Botero, Joseph A Tobias, Peter O Dunn, Hannah EA MacGregor, Dustin R Rubenstein, J Albert C Uy, Jason T Weir, Linda A Whittingham, and Rebecca J Safran. Sexual selection accelerates signal evolution during speciation in birds. In *Proc. R. Soc. B*, volume 280, page 20131065. The Royal Society, 2013.

[16] Emily A Ellis and Todd H Oakley. High rates of species accumulation in animals with bioluminescent courtship displays. *Current Biology*, 26(14):1916–1921, 2016.

[17] Janette Wenrick Boughman. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature*, 411(6840):944–948, 2001.

[18] Ken Kraaijeveld, Femmie JL Kraaijeveld-Smit, and Martine E Maan. Sexual selection and speciation: the comparative evidence revisited. *Biological Reviews*, 86(2):367–377, 2011.

[19] Edward H Morrow, Trevor E Pitcher, and Göran Arnqvist. No evidence that sexual selection is an 'engine of speciation'in birds. *Ecology Letters*, 6(3):228–234, 2003.

[20] Xavier Thilbert-Plante and AP Hendry. Five questions on ecological speciation addressed with individual-based simulations. *Journal of evolutionary biology*, 22(1):109–123, 2009.

[21] Josef C Uyeda, Stevan J Arnold, Paul A Hohenlohe, and Louise S Mead. Drift promotes speciation by sexual selection. *Evolution: International Journal of Organic Evolution*, 63(3):583–594, 2009.

[22] G Sander van Doorn, Pim Edelaar, and Franz J Weissing. On the origin of species by natural and sexual selection. *Science*, page 1181661, 2009.

[23] Rafał Dreżewski. The Agent-Based Model and Simulation of Sexual Selection and Pair Formation Mechanisms. *Entropy*, 20(5):342, 2018.

[24] James H Brown and Astrid Kodric-Brown. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology*, 58(2):445–449, 1977.

[25] Ilkka Hanski. Metapopulation dynamics. *Nature*, 396(6706):41–49, 1998.

[26] Paul F Doherty, Gabriele Sorci, J Andrew Royle, James E Hines, James D Nichols, and Thierry Boulinier. Sexual selection affects local extinction and turnover in bird communities. *Proceedings of the National Academy of Sciences*, 100(10):5858–5862, 2003.

[27] Susana Magallon and Michael J Sanderson. Absolute diversification rates in angiosperm clades. *Evolution*, 55(9):1762–1780, 2001.

[28] Jurriaan M De Vos, Lucas N Joppa, John L Gittleman, Patrick R Stephens, and Stuart L Pimm. Estimating the normal background rate of species extinction. *Conservation Biology*, 29(2):452–462, 2015.

[29] Erem Kazancıoğlu, Thomas J Near, Reinhold Hanel, and Peter C Wainwright. Influence of sexual selection and feeding functional morphology on diversification rate of parrotfishes (Scaridae). *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1672):3439–3446, 2009.

[30] Tim Janicke, Michael G Ritchie, Edward H Morrow, and Lucas Marie-Orleach. Sexual selection predicts species richness across the animal kingdom. *Proc. R. Soc. B*, 285(1878):20180173, 2018.

[31] Julia J Day, James A Cotton, and Timothy G Barraclough. Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PloS one*, 3(3):e1730, 2008.

[32] Peter M Todd and Geoffrey F Miller. Biodiversity through sexual selection. In *Artificial life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of living Systems*, volume 5, page 289. MIT Press, 1997.

[33] Susan E Masta and Wayne P Maddison. Sexual selection driving diversification in jumping spiders. *Proceedings of the National Academy of Sciences*, 99(7):4442–4447, 2002.

[34] Molly Schumer, Daniel L Powell, Pablo J Delclós, Mattie Squire, Rongfeng Cui, Peter Andolfatto, and Gil G Rosenthal. Assortative mating and persistent reproductive isolation in hybrids. *Proceedings of the National Academy of Sciences*, 114(41):10936–10941, 2017.

[35] Dean M Castillo, Amanda K Gibson, and Leonie C Moyle. Assortative mating and self-fertilization differ in their contributions to reinforcement, cascade speciation, and diversification. *Current zoology*, 62(2):169–181, 2016.

[36] LJ Campbell, ML Head, L Wilfert, and AGF Griffiths. An ecological role for assortative mating under infection? *Conservation Genetics*, 18(5):983–994, 2017.

[37] Steven R Beissinger. Ecological mechanisms of extinction. *Proceedings of the National Academy of Sciences*, 97(22):11688–11689, 2000.

[38] Marianne Elias, Rui Faria, Zachariah Gompert, and Andrew Hendry. Factors influencing progress toward ecological speciation. *International Journal of Ecology*, 2012, 2012.

[39] Steven M Stanley. Population size, extinction, and speciation: the fission effect in Neogene Bivalvia. *Paleobiology*, 12(01):89–110, 1986.

[40] Shaopeng Wang, Anping Chen, Jingyun Fang, and Stephen W Pacala. Speciation rates decline through time in individual-based models of speciation and extinction. *The American Naturalist*, 182(3):E83–E93, 2013.

[41] Robin Gras, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life*, 15(4):423–463, 2009.

[42] Abbas Golestani and Robin Gras. Multifractal phenomena in EcoSim, a large scale individual-based ecosystem simulation. In *ICAI (International Conference on Artificial Intelligence)*, pages 991–999, 2011.

[43] Abbas Golestani, Robin Gras, and Melania Cristescu. Speciation with gene flow in a heterogeneous virtual world: can physical obstacles accelerate speciation. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1740):3055–3064, 2012.

[44] Morteza Mashayekhi, Brian MacPherson, and Robin Gras. A machine learning approach to investigate the reasons behind species extinction. *Ecological Informatics*, 20:58–66, 2014.

[45] Marwa Khater, Dorian Murariu, and Robin Gras. Contemporary evolution and genetic change of prey as a response to predator removal. *Ecological Informatics*, 22:13–22, 2014.

[46] Maryam Karim Pour, Sourodeep Bhattacharjee, Robin Gras, and Ken Drouillard. The integration of an individual-based model into toxicokinetics to enhance ecological realism in evaluating population-level impacts of exposure to PCB. In *Complex Systems (WCCS), 2015 Third World Conference on*, pages 1–8. IEEE, 2015.

[47] Bart Kosko. Fuzzy cognitive maps. *International journal of man-machine studies*, 24(1):65–75, 1986.

[48] Adam Aspinall and Robin Gras. K-means clustering as a speciation mechanism within an individual-based evolving predator-prey ecosystem simulation. In *International Conference on Active Media Technology*, pages 318–329. Springer, 2010.

[49] Richard M Sibly, Volker Grimm, Benjamin T Martin, Alice SA Johnston, Katarzyna Kułakowska, Christopher J Topping, Peter Calow, Jacob Nabe-Nielsen, Pernille Thorbek, and Donald L DeAngelis. Representing the acquisition and use of energy by individuals in agent-based models of animal populations. *Methods in Ecology and Evolution*, 4(2):151–161, 2013.

[50] Nazia Hashim, Sonia Naqvi, Majida Khanam, and Hassan Fatima Jafry. Primiparity as an intrapartum obstetric risk factor. *JPMA-Journal of the Pakistan Medical Association*, 62(7):694, 2012.

[51] Joachim Künkele. Does primiparity affect the efficiency of converting energy to offspring production in the guinea-pig? *Canadian Journal of Zoology*, 78(2):300–306, 2000.

[52] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[53] Sreerama K Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4):345–389, 1998.

[54] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[55] JR Quinlan. C4. 5: Programs for Empirical Learning Morgan Kaufmann. *San Francisco, CA*, 1993.

[56] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[57] Mark A Hall and Lloyd A Smith. Practical feature subset selection for machine learning. 1998.

[58] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[59] Shumeet Baluja. Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, 1994.

[60] Heinz Mühlenbein and Gerhard Paass. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*, pages 178–187. Springer, 1996.

[61] Pedro Larrañaga and Jose A Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.

[62] Martin Pelikan, David E Goldberg, and Fernando G Lobo. A survey of optimization by building and using probabilistic models. *Computational optimization and applications*, 21(1):5–20, 2002.

[63] Iñaki Inza, Pedro Larrañaga, Ramón Etxeberria, and Basilio Sierra. Feature subset selection by Bayesian network-based optimization. *Artificial intelligence*, 123(1-2):157–184, 2000.

[64] Marwa Khater, Elham Salehi, and Robin Gras. Correlation between genetic diversity and fitness in a predator-prey ecosystem simulation. In *Australasian Joint Conference on Artificial Intelligence*, pages 422–431. Springer, 2011.

[65] Elham Salehi and Robin Gras. Efficient EDA for large opimization problems via constraining the search space of models. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 73–74. ACM, 2011.

[66] Sourodeep Bhattacharjee and Robin Gras. Estimation of Distribution using Population Queue based Variational Autoencoders. *In Press, IEEE CONGRESS ON EVOLUTIONARY COMPUTATION*, 2019.

[67] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[68] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic back-propagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[69] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[70] Yasmin Latour, Marco Perriat-Sanguinet, Pierre Caminade, Pierre Boursot, Carole M Smadja, and Guila Ganem. Sexual selection against natural hybrids may contribute to reinforcement in a house mouse hybrid zone. *Proc. R. Soc. B*, 281(1776):20132733, 2014.

[71] M Tobler, R Riesch, CM Tobler, T Schulz-Mirbach, and M Plath. Natural and sexual selection against immigrants maintains differentiation among micro-allopatric populations. *Journal of evolutionary biology*, 22(11):2298–2304, 2009.

[72] Richard Abbott, Dirk Albach, Stephen Ansell, Jan W Arntzen, Stuart JE Baird, Nicolas Bierne, Jenny Boughman, Alan Brelsford, C Alex Buerkle, Richard

Buggs, et al. Hybridization and speciation. *Journal of evolutionary biology*, 26(2):229–246, 2013.

[73] Peter R Grant and B Rosemary Grant. Hybridization of bird species. *Science*, 256(5054):193–197, 1992.

[74] Michael L Arnold and Noland H Martin. Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*, 25(9):530–536, 2010.

[75] Yuan Yuan, Stephen T Buckland, Phil J Harrison, Sergey Foss, and Alison Johnston. Using species proportions to quantify turnover in biodiversity. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(2):363–381, 2016.

[76] Melanie J Monroe and Folmer Bokma. Do speciation rates drive rates of body size evolution in mammals? *The American Naturalist*, 174(6):912–918, 2009.

[77] R Brian Langerhans. Trade-off between steady and unsteady swimming underlies predator-driven divergence in Gambusia affinis. *Journal of evolutionary biology*, 22(5):1057–1075, 2009.

[78] R Brian Langerhans, Matthew E Gifford, and Everton O Joseph. Ecological speciation in Gambusia fishes. *Evolution*, 61(9):2056–2074, 2007.

[79] P Raia, F Carotenuto, A Mondanaro, S Castiglione, F Passaro, F Saggese, M Melchionna, C Serio, L Alessio, D Silvestro, et al. Progress to extinction: increased specialisation causes the demise of animal clades. *Scientific Reports*, 6, 2016.

[80] Marcel Cardillo and Lindell Bromham. Body size and risk of extinction in Australian mammals. *Conservation biology*, 15(5):1435–1440, 2001.

[81] German Forero-Medina, Marcus Vinícius Vieira, Carlos Eduardo de Viveiros Grelle, and Paulo Jose Almeida. Body size and extinction risk in Brazilian carnivores. *Biota Neotropica*, 9(2):45–49, 2009.

Chapter 5

Animal communication of fear and safety related to foraging behavior and fitness

5.1 Introduction

There are a number of empirical studies that have demonstrated the influence of animal communication upon behavior. Vibrational communication in insects influences mating behavior and sexual selection ([1]. Authors in [2] have reported the combination of auditory and visual courtship signals to have a stronger influence on courtship behavior in female pigeons (cooing, circling and tail spreading) than either modality acting singularly. Nobel laureate Karl von Frisch found that honey bees (genus apus) initiate waggle dances, thereby communicating to conspecifics the location of food, resulting in influence on foraging behavior (von Frisch, 1967). Along the same lines, [3] reported that bumblebees (Bombus terrestris) communicate the existence of honey spots to conspecifics after they unload by grooming and fanning their wings,

which results in recruits for foraging activity. Also, there exists evidence of the role of alarm signalling in the communication of urgency for escape from predators in the bird Sericornis frontalis [4]. The authors discovered that multi-modal alarm signals were more likely to result in escape behavior than unimodal signals [4]. There is also a body of literature, described below, that investigates the influence of alarm communication plays on foraging behavior, which is the focal point of this article. There is a high level of corroboration within empirical studies supporting the claim that predator alarm cues and presence of predation decreases foraging activity in prey. For instance, as reported by [5], hard clams, Mercenaria mercenaria reduce their feeding behavior in response to cues from injured conspecifics. Further, chemical alarm cues generated from closely related heterospecifics (but not from distantly related heterospecifics) were associated with reduced foraging behavior in a coral reef fish, Amphiprion percula [6]. [7] observed that a species of termite workers, Coptotermes acinaciformi reduced their foraging behavior in response to vibratory alarm signals elicited by soldiers. [8] reported that crab foraging behavior lessened for the species Panopeus spp. in the presence of cues from predators. Further, [9] found that snails of the species Physa gyrina sought refuge and were less active in response to injured snail cues as a result of predation. [10] reported that convict cichlid fish, Archocentrus nigrofactiatus, exibits an altered foraging response in the face of damage release cues from conspecifics (See also [11] ; [12] for similar results).

Opposition to the view that increased predation and alarm cues result in lower foraging activity in prey, [13] is the predation risk allocation hypothesis. The central departure involved forecasts an increase in foraging activity by prey in response to continued predation and alarm cues in order to secure nourishment over the course of time. This hypothesis involves a method of cost-benefit analysis where, over the course of time, the cost of vulnerability to predation is outweighed by the benefits of finding

food to survive. As [14] observed, the predation risk allocation hypothesis entails two predictions: 1) initial predation and predator alarm cues result in a decrease in foraging activity, although 2) as exposure to predation and alarm cues remain high over a longer period of time, prey begins to engage in risky foraging behavior given the need for nourishment and energy. As [15] noted, a number of studies seem to corroborate the predation risk allocation hypothesis including [16], [17] along with [18]. In all of the aforementioned studies, prey demonstrated higher foraging activity in the presence of sustained high levels of predation and alarm cues. In contrast, authors in [15] argued the findings of [11] and [12] demonstrated a decrease in foraging behavior in the presence of predation (which is consistent with the findings we presented in the first paragraph), present evidence for the predation risk allocation hypothesis to still remain unanswered. [14] also expressed a measure of skepticism regarding the predation risk allocation hypothesis. An important goal of our study is to enhance understanding of this debate. In particular, we seek to test the following hypotheses in our study:

H1: Communication of fear related to predation is associated with a lessening of foraging behavior relative to the absence of communication.

H2 (predation risk allocation hypothesis): Communication of fear related to predation is associated with an increased level of foraging behavior over time relative to the absence of communication. These hypotheses are not necessarily mutually exclusive as fear communication results in an initial decrease of foraging activity. Yet, over a period of time, some animals may become habituated to fear communication and take risks in order to increase inclusive fitness. This notion provides a relative basis for the third hypothesis examined by our study involving the claim that habituation to alarm communication augments inclusive fitness. With respect to H2, the idea that alarm communication augments fitness after habituation to alarm occurs aligns

with the expectation that an increase in individual fitness would occur over time due to the increase of foraging behavior despite alarm signals. The basis for this hypothesis assumes that nutrition is required for reproduction. If this expectation is met, it would appear to corroborate H2. Conversely, a fourth hypothesis to be considered in this investigation centers on a decrease in foraging activity in response to alarm communication that increases individual fitness since an animal cannot reproduce if it is killed by a predator. Confirmation of this hypothesis would appear to corroborate H1, since it is plausible for alarm communication to result in a decrease in foraging behavior, thereby increasing individual fitness. Thus, an additional question that we will be investigating in this paper is the connection between alarm communication and its effect on foraging behavior and individual fitness. Fitness of individuals in EcoSim is measured as the sum of the number of children and grandchildren produced by an individual. This characterization of fitness was developed in brommer2004single and later in [19].

In order to address the issue of this possible relationship, we will investigate the following additional hypotheses:

H3: Alarm communication augments individual fitness in animal species when there is habituation to alarm communication,

H4: Alarm communication augments individual fitness in animal species when this communication reduces foraging activity.

From an overview of the literature, habituation to alarm cues suggested by the risk allocation hypothesis appears to decrease fitness, which challenges H3. Authors in [20] described habituation to alarm pheromones over 3 generations as a consequence of nutritional requirements that resulted in lower survival (presumably due to being killed by predators) and fecundity in aphids. Higher fitness was exhibited by aphids that did not habituate to perception of alarm signals, which corroborates H4. More-

over, [21] reported that cichlid fish (P. cichlid fish) reduced activity in response to alarm cues from conspecifics, thus enhancing survival and hence fitness, which again appears to refute H3 and corroborate H4. Moreover, additional published support exists for H4 - the hypothesis that alarm communication that decreases foraging behavior and results in increased inclusive fitness in a variety of animal species. Authors in [22] reported that tetra fish (Hemigrammus erythrozonus) responded to subthreshold chemical cues from conspecifics by increasing their vigilance against predators, which the authors designated as a 'fitness-related' behaviour. An earlier study by [23] studied 6 species of monkeys where individuals elicited alarm signals in the presence of predators such as leopards. The authors concluded that these signals served the primary function of warning conspecifics rather than alerting predators, which they argued to provide kinship advantages and augment inclusive fitness [23]. There also appears to be some animal species that use alarm signals to alert predators rather than conspecifics. Various rodent species as discussed in [24] and [25] exhibited this modality, which was found to increase indirect fitness. We shall conduct individual-based modeling computer simulations in our study in order to test hypotheses H1 through H4. The significance of our study is that it will clarify the relationship between fear communication and foraging behavior and discern how this impacts fitness across both specific and general animal species.

5.2 Methods

5.2.1 EcoSim

EcoSim ([26] ; [27]) is a simulator of a virtual individual based ecosystem designed to simulate the behavior of predators and prey individuals in a dynamic tri-trophic environment consisting of 1000 x 1000 cells (EcoSim source code (in C++) can be

obtained from the repositories at https://github.com/EcoSimIBM). EcoSim serves the purpose of studying complex ecological problems when time and cost issues make field studies impractical. EcoSim has been used to investigate a wide range of ecological properties such as speciation rate [28], extinction of species [27], evolution of prey individuals under the stress of predator pressure [29], ecological risk assessment of Polychlorinated biphenyls [30], and effect of different mate selection strategies on extinction and speciation [31].

EcoSim uses a fuzzy cognitive map (FCM; [32]) to model behavior of individuals. These FCMs are coded in the genome of the individuals and are subject to evolutionary processes through crossover and mutation for both predators and prey. Another essential feature within EcoSim pertaining to the present paper is the capacity of fitness of individuals to exist as an emergent property of the system; that is the individuals are not programmed to attain higher fitness. Instead, we define the fitness of an individual as the sum of children and grandchildren produced by that individual. This definition of fitness was developed in [33] and in [19], and it was discussed in [34].

Each cell in EcoSim may contain prey, predators, and grass. Grass serves as the source of energy for prey, while the predators hunt and consume prey or meat (previously killed prey). The simulation goes through a discrete succession of steps, henceforth referred to as a timestep. During each such timestep, each predator and prey perform one unique action based on perceived environmental factors and internal state (facilitated by FCMs). Examples of potential actions taken by prey include escaping from predators, searching for food, socializing with other prey, exploring the virtual world of EcoSim, resting, eating and reproducing.

To achieve such actions, all individuals are required to spend a certain amount of energy based on the bioenergetic model and physical factors of the particular individual. Speed, maximum speed, vision range, and reproductive age, (repAge), are

examples of factors involved. The exact equation for energy expenditure of prey individuals is given by equation 5.1. NbArcs is a measure of the complexity of the brain based on the number of edges in the FCM of an individual. Vision is the distance (number of cells) for which individuals can see. Defense is the measure of capability of an individual to protect itself from predator attacks, while coopDefense (Cooperative Defense) is the ability of an individual to protect other local prey from predators.

$$\begin{aligned} Energy_{prey} = 0.8 \times max((nbArcs - threshold)^{0.75}, 1)) + \frac{Strength \times (Speed)^2}{10000} \\ + \left(\frac{MaxEnergy}{5}\right)^{0.75} \\ +(Vision \times 5)^{0.75} + (Defense \times 5)^{0.75} + (CoopDefense \times 75)^{0.75} \\ +(Max(0, 8 - repAge))^{2.3} \end{aligned} \tag{5.1}$$

In EcoSim the genome of an individual consists of two sections : the physical genome representing the physical features and the behavioral genome coding for the behavioral model of the individual. The genome of an individual remains constant during its lifetime. The behavioral model of an individual is modeled using a FCM, which is a directed graph consisting of nodes and edges (see figure 5.1). Due to the evolutionary process modeled by EcoSim, each individual has a unique FCM. The nodes represent various concepts, while the edges represent various relationships between concepts. These concepts can be of three types: sensory concepts (sensing of distance from predator, perception of own energy level, etc.), internal concepts (such as fear, satisfaction, disgust, etc.) and action concepts (shown in figure 5.2). Each edge has a weight associated with it, which represents the degree of influence of a source concept upon a destination concept. These edge weights can be negative or

positive and are based on the type of influence one concept has upon another. Hence, a positive edge weight means the source has an excitatory influence on the receiving concept, while a negative edge means the source edge has an inhibitory effect on the receiving concept.

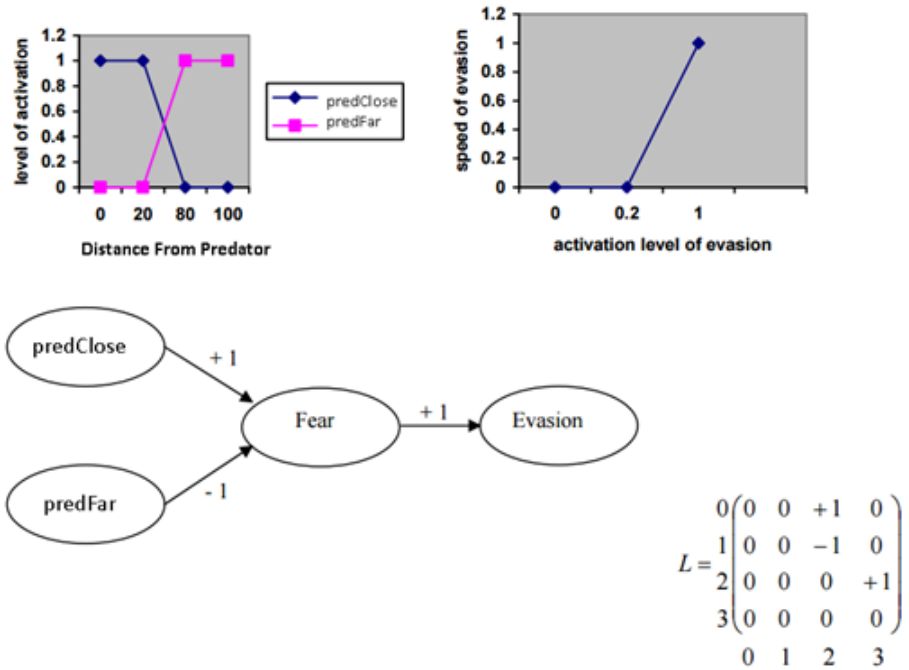
5.2.2 Experiments

To test our hypotheses, ten independent runs of EcoSim were conducted, where all prey individuals possessed the ability to communicate fear (Communication Runs); ten other independent runs were also conducted where individuals did not have communication capabilities (Non-Communication Runs). Each set of runs was executed to 36,000 time steps.

To model communication between individuals, two new sensory concepts were created for prey individuals: "Communicated Fear" (CommDanger with values between 0 and 1) and "Communicated Safety" (CommSafety –with values between 0 and 1) (see figure 5.3). The input to the CommDanger node of prey is the average activation level of fear of all prey (say f) within 3x3 cells centered around the given prey. In contrast, the input to the CommSafety concept is 1-CommDanger (which is an indicator of safety of the environment).

Figure 5.4 shows how the receiver obtains the extra information related to fear. In a scenario where receiver prey B is unable to sense a predator directly due to too large of a distance, perception of distant predators is still possible through the fear level of neighbors, which allows prey to act accordingly. It is to be noted that the information about the fear level of each individual (CommDanger) is transmited to all individuals in a 3x3 cells block irrespective of which prey species the individuals belong.

Initially, edges are not assigned from CommDanger and CommSafety to other



$$L = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{array}{c} \\ \\ \\ \\ \end{array}$$
$$\begin{array}{cccc} 0 & 1 & 2 & 3 \end{array}$$

Figure 5.1: An example of a simple FCM in which activation of predClose (proximal predator) and predFar (distant predator) is given by fuzzification of these concepts, depending on the distance of prey from predator. In the fuzzification process, the real value of the sensory concept (consider a predator 5 cells away) is converted to a fuzzy value (a decimal number from 0 to 1). The speed at which prey evade is given by defuzzificaton of the evasion concept, where the reverse of fuzzification happens - the fuzzy value is converted to a real scalar value. The L matrix is an nxn matrix showing the influence of one concept upon another; where 0 denotes predClose, 1 denotes predFar, 2 represents fear, and 3 represents evasion. Activation levels of motor concepts in EcoSim dictate choice of action for an individual and the defuzzification of the activation level provides the intensity of the action. For example, if the evasion concept is activated, the defuzzification of the evasion concept gives the speed of evasion. [26]

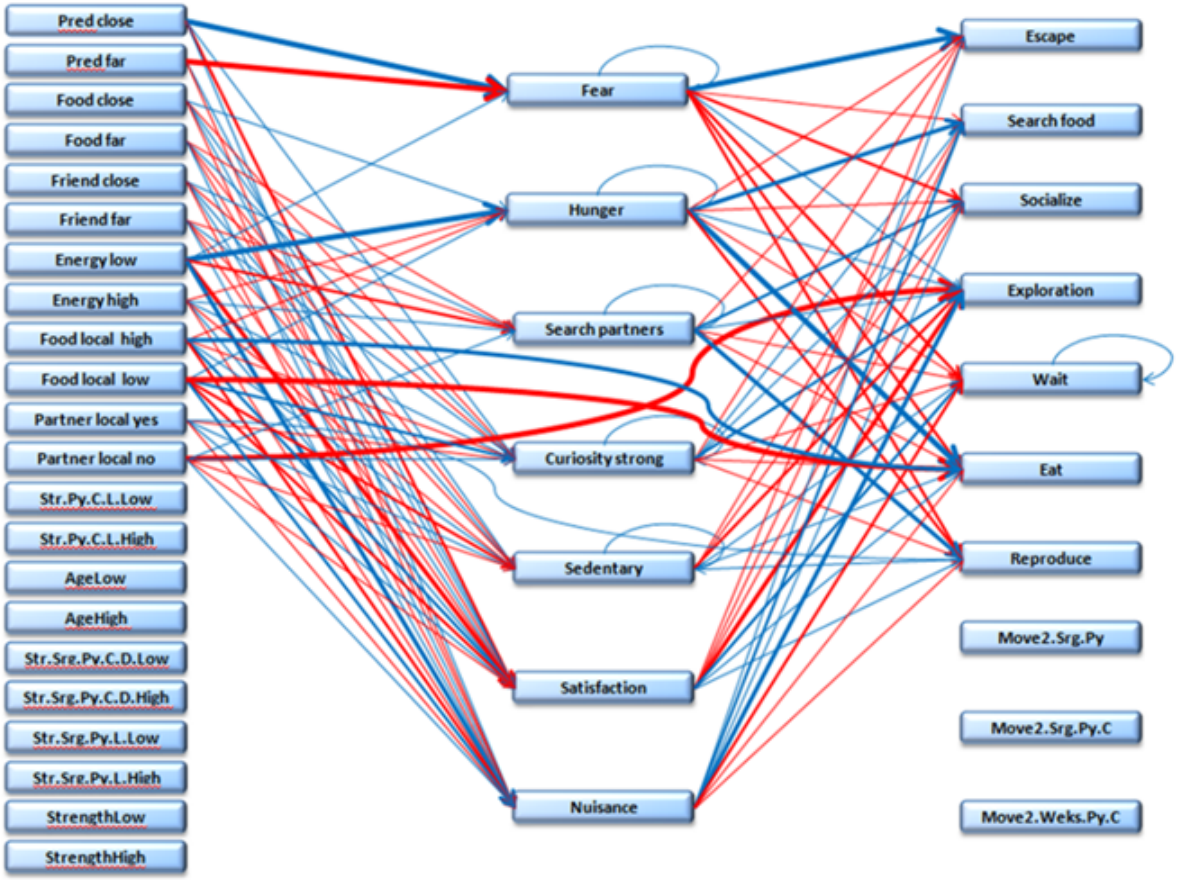


Figure 5.2: FCM of a typical prey individual. The left column of nodes contains the sensory concepts; the middle column of nodes contains the internal concepts, and the right column of nodes contain the motor concepts. The red edges denote a negative edge and a blue line denotes a positive edge. The thickness of the lines indicates the weight of the edge.

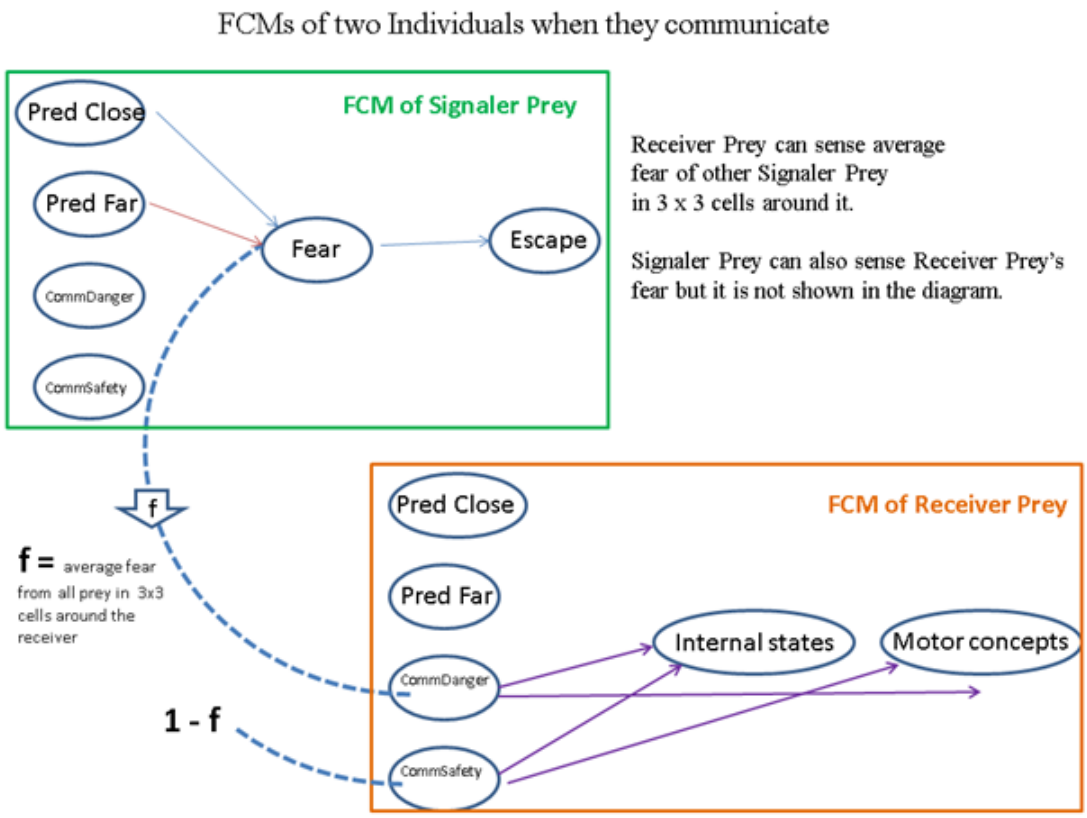


Figure 5.3: An illustration of Communication-related to Danger and Safety shows the sharing of fear between two individuals near one another.

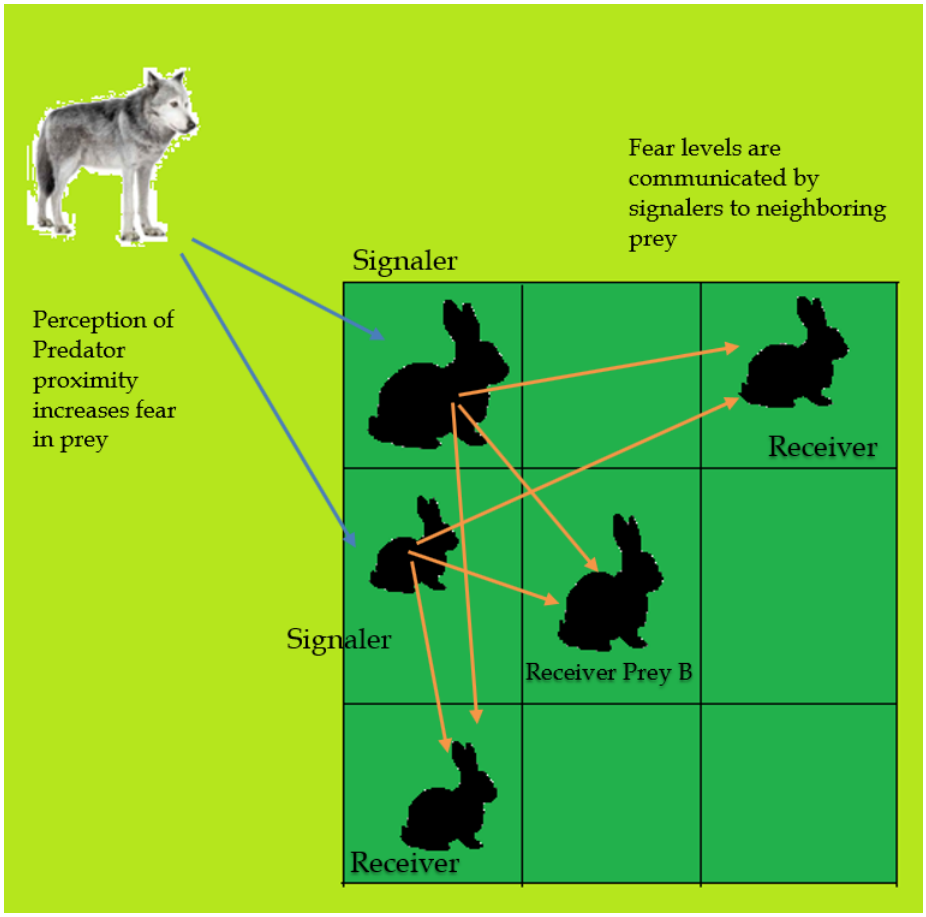


Figure 5.4: Communication of Fear from Signaler to Receiver. Receiver prey received the extra information related to fear, even when the receiver is unable to sense a predator directly due to a greater distance

nodes. The way individuals use the perceived information depends on the emergence of new edges between the two new concepts and other internal and motor concepts. The emergence of new edges is the result of evolutionary processes within EcoSim.

To evaluate the evolutionary value of this communication system, a cost has been added for the usage of the CommDanger/CommSafety node information. The cost is calculated as 0.09 % of maximum energy of an individual at each time step. The addition of this cost implies that an individual expends energy as soon as an edge develops between the CommDanger or the CommSafety nodes to any other node in the individual's FCM. This models the cost of having a new perception mechanism. This cost supercedes the cost associated with each edge, including the ones connected to the CommDanger/CommSafety nodes.

At this stage, it is important to distinguish the experimental method used in our article from the experimental method employed in the recent simulation study conducted by [35] as there is a key difference between our EcoSim simulations regarding prey individuals and the simulations conducted by the aforementioned authors. The simulations performed by authors in [35] involve 3 types of runs with prey individuals and additionally employ 3 distinct foraging strategies - control (like our non-communication prey), risk-averse, and risk-taking. Our simulations assign no distinction between risk-averse vs. risk-taking communities and employ only 2 types of runs: communication and non-communication. In communication runs, communities averse to risk and not averse to risk were *emergent* in our simulations over time rather than occur as parameters for initialization.

5.2.3 Machine Learning

Decision Trees

Decision tree algorithms yield a set of hierarchical rules based on features of data that aid in dividing datasets recursively into various classes [36]. Hence, decision trees shed light on significant distinctive properties of data. Each path in a decision tree, from the root to a leaf node, corresponds to a rule that can be used to predict the dependent feature based on independent features. Thus, the rule governing whether a data point corresponds to a communication run or a non-communication run can be used to predict the population density in the world or the level of predation fear in the world, for example.

In univariate decision trees, the hierarchical split of data instances is based on a single feature at any given node and takes of the form of: $x < c$ or $x > c$, with x representing the feature and c representing a threshold value. At a given node along a pass, all the current data instances for which the decision feature x is less than a certain constant are sent to the left downside branch, while the other instances are sent to the right branch. This process generates a split of the current set of instances into two subsets. Each of those subsets of instances undergoes further splits until the dependent feature can be classified with some degree of accuracy.

At the level of a node, an exhaustive search is used to find the independent feature x and the constant c, where the objective is set to optimize some measure of node impurity, such as entropy [37]. This measure of entropy takes a value between zero and one depending on how the instances are split by the node. If all the instances fall into one category, entropy takes a value of 0; entropy takes a value of one when the instances are equally divided into two classes.

In this paper we have used a specific decision tree algorithm known as C4.5 [38]

which uses the information gain ratio as an entropy measurement to identify how well a decision variable can split data entering the decision feature.

Variational Autoencoder EDA as a Wrapper Method for Feature Selection

Machine learning algorithms extrapolate knowledge from massive datasets. The reliability and interpretability of knowledge gained through machine learning analysis of datasets can be affected by the presence of irrelevant, redundant, and superfluous information (features) present in such datasets [39]. Feature selection is a class of machine learning algorithms used to detect and expunge datasets of extraneous features before the predictive model is built. To illustrate, consider a dataset that contains samples described by 10 features, but only 3 of these features are useful in classifying the dataset into one category or another. Feature selection algorithms could be employed to remove the other 7 features so that a more concise and efficient classifier could be built.

Wrapper methods are a type of feature selection algorithms where the model building step (such as generating a decision tree) is integrated within the feature selection step [40]. This differs from the filter method in which feature selection is performed as a preprocessing step before the model is built. Wrapper methods define a combinatorial optimization search procedure in the complete feature space that returns subsets of features to be evaluated based on the predictive model. In the evaluation phase, these subsets of features are used to train and test a specific classification model (a C4.5 decision tree, for example), and the accuracy of the decision model generated from this subset is used to judge the worthiness of the feature subset. Thus, to search the feature space for the best subset, the search is wrapped around the classification model.

Estimation of distribution algorithms (EDA) [41–44] are metaheuristics that aid

in combinatorial optimization. EDA maintains a population of promising candidate solutions. In a selection step, the candidate solutions with the best fitness are picked from the population and thereafter a probabilistic generative model is built from these promising candidates to generate the next generation of promising candidates. Fitness of the candidate solutions is defined by some problem specific criteria that evaluates the candidate as a whole, based on the values of each constituent feature in the candidate solution. In our methodology, the fitness function is the predictive models accuracy. This process repeats itself until the optimal solution is obtained or the current best solution has a fitness considered to be sufficient. The process to determine the exact combination of assignments to variables that lead to a higher fitness is complex. EDAs have been used successfully as a wrapper method for classification tasks in [45] and [46]. In [46] Constrained Model Search Space Bayesian Optimization Algorithm (CMSS-BOA) [47] was used as a wrapper method with C4.5 Decision trees. When EDAs are used as the Wrapper Method, the fitness is evaluated based on the predictive accuracy of the classifier.

Variational Autoencoder EDA with Population Queue (VAE-EDA-Q) [48] uses variational autoencoders [49] [50] to model the population. This method demonstrated efficacy in exploring latent continuous search spaces. VAE-EDA-Q is able to explore continuous latent space representation of the features which allows randomized sampling and smooth interpolation over the manifold of high-quality solutions in the feature space.

In this paper, we utilized a wrapper method combining C4.5 algorithm with VAE-EDA-Q to discover the distinguishing characteristics of communication and non-communication runs. The capacity of VAE-EDA-Q to act as a combinatorial optimizer to select the best subset of features yields optimal accuracy for C4.5 algorithm.

5.3 Results and Discussion

5.3.1 The effects of alarm communication on foraging behavior (Comparing communication and non-communication runs)

As discussed in the introduction, two hypotheses examined by our simulation include:

H1: Communication of fear relating to predation is associated with a lessening of foraging behavior relative to the absence of communication.

H2 (predation risk allocation hypothesis): Communication of fear relating to predation is associated with an increased level of foraging behavior over time relative to the absence of communication.

The average of all runs appear to corroborate H1 and refute H2, as can be seen from figure 5.5 . Over time (by the 20,000th time step), communication relating to the presence of predators is associated with an overall decrease in foraging behavior, whereas the absence of communication is associated with an overall increase in foraging behavior.

To determine the statistical significance of the observed differences, Kruskal Wallis One Way ANOVA tests were conducted individually on the test parameters. Before non-parametric tests were conducted, averages were performed for groups of 30 data points in each window (for thirty consecutive time steps) and omission of the subsequent 20 timesteps was completed to remove inherent temporal continuity in the results. Therefore, the data points obtained were for averages of timesteps 1- 30, 50 – 80, 100 – 130, etc. Skipping 20 timesteps in between yields discrete samples on which ANOVA can be applied, as the direct causal dependence from one sample point to the next is removed.

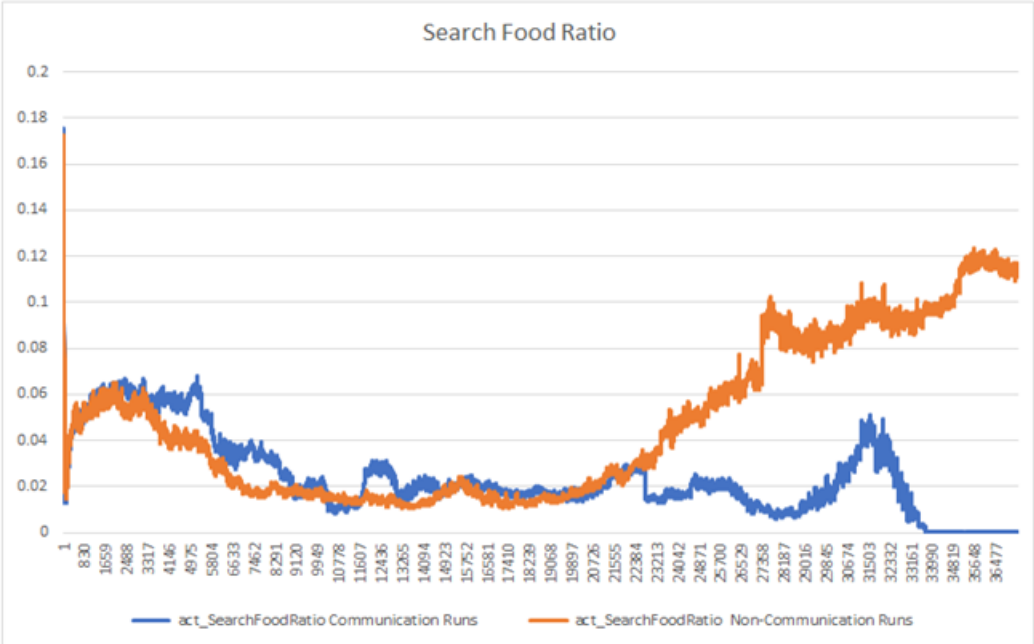


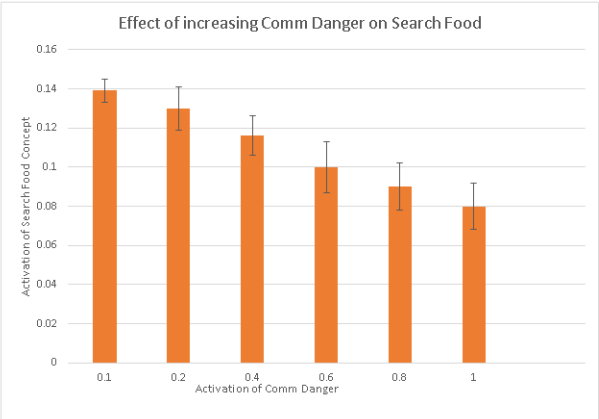
Figure 5.5: Comparing Search Food Ratio between Communication and Non-Communication Runs

The dependent variable is the Search food ratio, which is defined as the ratio of the number of individuals that opted for the search food action to the total number of individuals for a given time step. ANOVA tests were also performed to treat each run as an independent sample, Search Food ratio was averaged for all time steps occurring after 20000th generation. For ten runs of Communication and ten runs of Non-Communication, 20 independent and normal data points were chosen for ANOVA testing to ascertain whether the independent variable (Communication Run/ Non-communication Run) could explain the differences in Search Food ratio. The differences were found to be statistically significant from ANOVA experiments analyses (p-value < 0.0001, alpha $= 0.05$).

These results concur with the empirical findings of [5]; [6]; [7]; [8] and [9]. They are in disagreement, however, with other empirical studies such as [16]; [17] and [18] which corroborate H2, the risk allocation hypothesis.

An aparent issue with respect to our data is the significant divergence at roughly 20,000 time steps for communication vs. non-communication runs displayed by our graphs for the Search Food Ratio. Potentially, this divergence could explained by the mechanism of *evolution* of foraging behavior in both communication and non-communication species rather than *phenotypic plasticity*. Evolutionary change generally occurs multi-generationally, while phenotypic change is relative to the existing genotype [51–53]. This is not to presume that evolution takes thousands or millions of years in all cases, as is observed by [54] in his work on eco-evolutionary dynamics, as well as in [29]. [54] noted that Darwin's finches evolved beak sizes suited to their environment over a relatively short period of time. Similar findings were observed by [55] where evolutionary changes in guppies took place only after 30 – 60 generations. Along these lines, it is important to note that one time-step in EcoSim is not equivalent to a generation in biological terms. One generation can be represented by multiple time steps, since within EcoSim, a time step only involves one decision, one action, and any changes regarding species status. Thus, 20,000 time steps do not equal 20,000 generations. In fact, one generation, which is linked to the age of reproduction, equals 6 to 8 time steps [56].

At this point, it could be argued inadvertent conflation of two types of communication relating to predation occurred – Comm Danger and Comm Safety. Further analysis, however, suggests that these two types of communication are consistent in their effects on foraging behavior overall. As can be seen in Figure 5.6, Comm Danger is strongly correlated with a decrease in foraging behavior (Correlation coefficient = -0.9948) and Comm Safety is strongly correlated with an increase in foraging behavior (Correlation coefficient = 0.9967) – averaged across all runs. This is expected relative to H1.



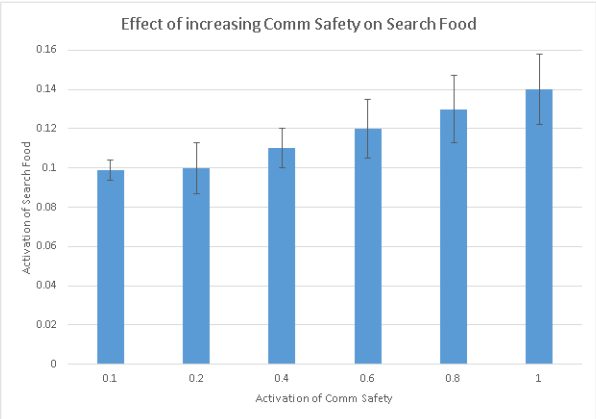


Figure 5.6: Comparing Search Food Ratio between Communication and Non-Communication Runs

5.3.2 Difference in effect of communication on Search Food across various Communication runs

The number of individuals using a certain FCM edge within a range of weight values is recorded to assess and associate a level of significance to the usage of an edge within the whole population (figure 5.3). Figure 5.7 illustrates the methods for which communication nodes can be used in the population. The y-axis of the graph shows the number of prey individuals using a particular communication edge, at a particular timestep for one run in our simulation batch, while the x-axis gives the names of edges connecting Communicated Danger or Communicated Safety to other nodes in the FCM.

To quantify the impact of the communication nodes, a combined usage index of the usage data was calculated to find the mean usage of an edge in the population. The combined usage index is calculated by multiplying the number of individuals using a particular edge, within a weight value range, by the higher edge weight value in that range, and adding up the values for each range for a given link (i.e. the link between CommDanger and SearchFood (See Supplementary Material A)).

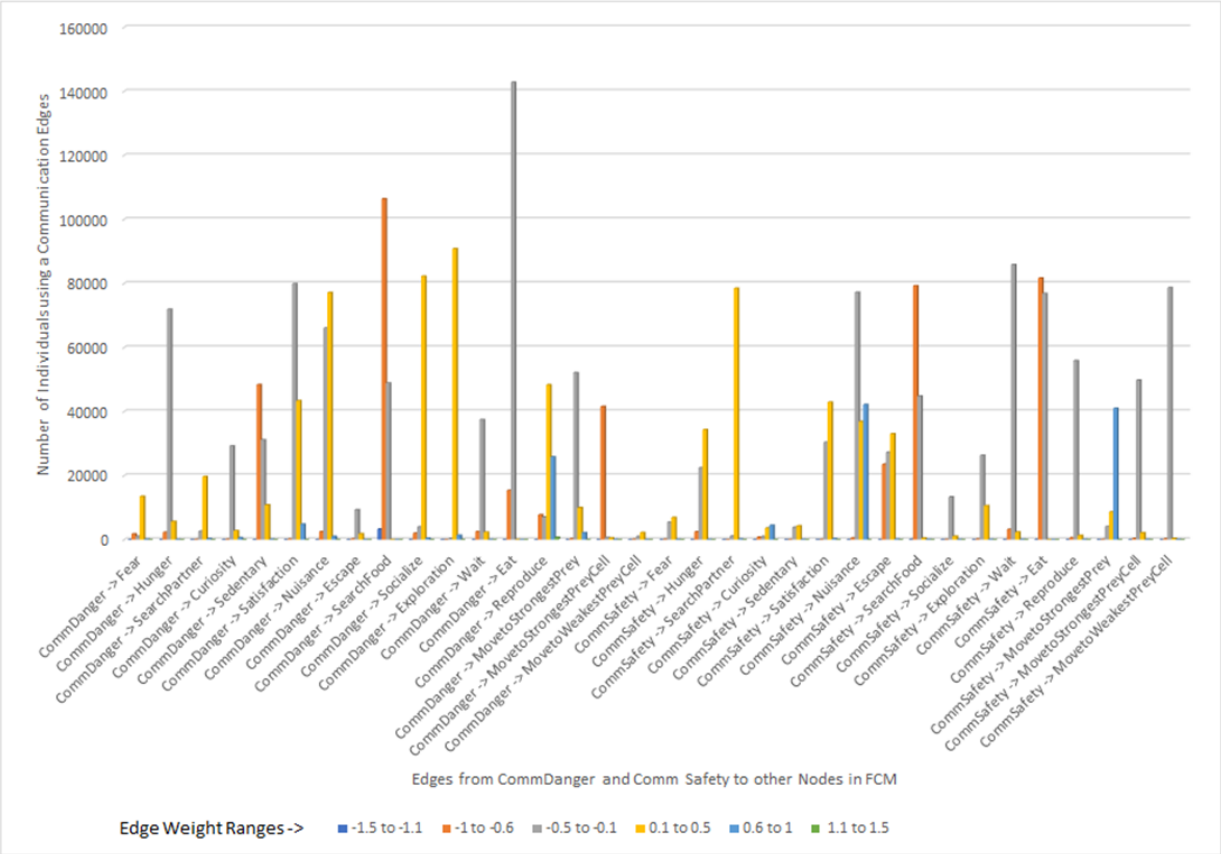


Figure 5.7: Number of individuals using links from Communication nodes to other nodes for Run 8 at 34000th timestep.

Using the concept of the combined usage index, the net effect of CommDanger and CommSafety on Search Food in individual communication runs can be measured. In figure 5.8 below, the majority of the individual runs displayed edge values linking communication with search for food as negative and, hence, inhibitory when averaged across individuals. However, some of the individual runs (particularly 3, 4 and 5) displayed edge values that were either positive, negligible or only slightly negative. This suggests that in some circumstances, fear-related communication has the potential to increase (or at least not diminish) foraging behavior when the benefits of securing nourishment outweigh the risks associated with predation. This result aids in the corroboration of the Risk-Allocation hypothesis, H2, outlined above. That is, there exists, in some cases, habituation to alarm communication over time if the risks associated with predation are outweighed by the fitness-related benefits of finding food.

Although the combined usage index provides some information about the direct impact of the communication nodes on a given behavior, such as search-food, it does not capture the potential indirect influence communication nodes may have. For example, in addition to the direct edge between CommDanger to Search-food, there are also indirect connections between these two concepts, such as CommDanger $->$ Fear $->$ Search-food. The communication nodes can exert influence on the behavior via these indirect connections as well, and these indirect influences have potential significance. To further examine the overall effects of CommDanger and CommSafety on Search food that accounts for both direct and indirect connections, we developed a simulation tool to analyze the FCM of each individual outside of the dynamic environment of EcoSim (see Figure 5.9). Using the tool, the input of the CommDanger (sensory node) was varied while all other sensory input was fixed, and variation within Search Food was measured under these conditions. Similarly, the input of the Comm-

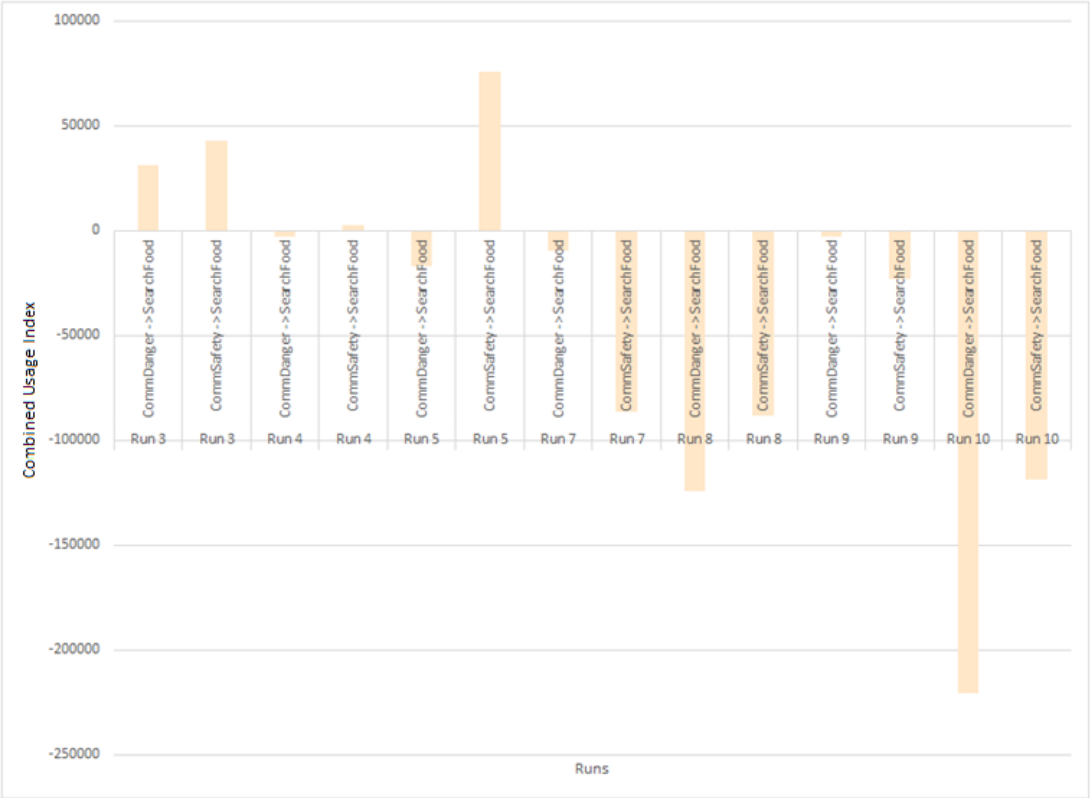


Figure 5.8: Individual run impacts of Communicated Danger and Communicated Safety on Search Food Node at 34000th timestep.

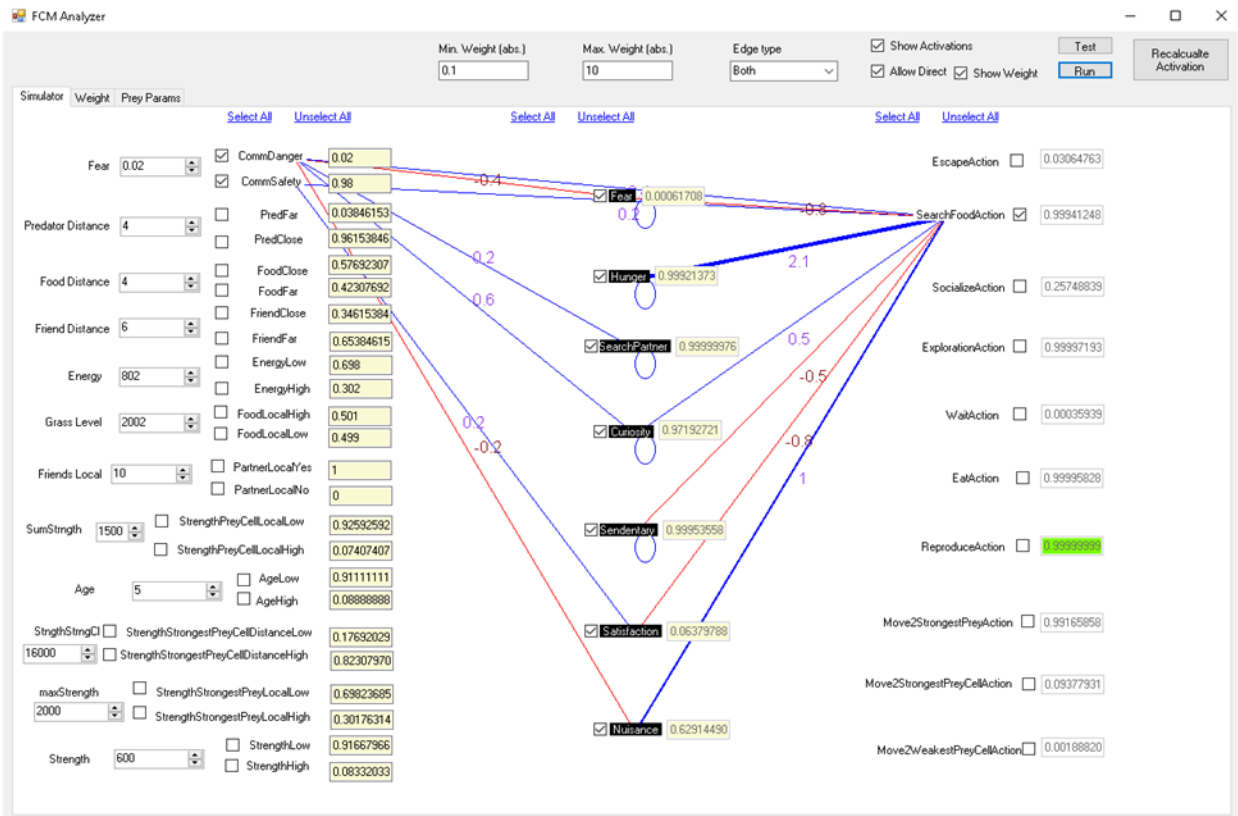


Figure 5.9: FCM Analyzer showing the FCM network of one prey individual and the connections from Communication Node to various internal nodes and finally stopping at Search Food.

Safety was varied while all other sensory input was fixed, and variation in Search Food was again measured in this similar fashion. FCM values (edge weights and connections) of 50 random prey individuals were extracted from the end of each of the 10 runs. Upon them, the experiment to find the independent effects of CommDanger and CommSafety on Search-Food ratio was performed. For each run, the activation of search food was measured for each of the 50 prey individuals and the average of the activation of search food was reported. Other sensory node values were set constant and derived from the average value for the set of 50 individuals from each run.

In our experiments we discovered that each communication run assumed one of two behaviors (see figure 5.10). In one set of runs, the impetus to search for food

increased with increase of communication. While in another group of runs, the search food behavior decreased with an increase in communication. To better understand the independent effects of communicated danger and communicated safety on search food in these two groups of runs, the same tool was used to measure the extent of the effect of Communication on search food in two groups of communication runs. Effects resulted from 3 runs in each group, where 100 individuals from each run were averaged for the 3 runs in each group.

A more fine-grained analysis of the data, where the independent effect of Comm Danger and CommSafety upon Search food across all runs was not averaged but instead considered individually, revealed a more nuanced result. Within the specific runs of 7,8, and 10, H1 which states alarm communication to result in decreased foraging was corroborated, Within the specific runs of 3,4 and 5, H2, the risk allocation hypothesis, was corroborated. Figure 5.10 shows the independent effects of CommDanger and CommSafety on Search Food for the two sets of runs, where 100 individuals from each run were averaged over 3 runs in each set. Two significant findings can be drawn from this data. First, the effects of communication appear to be primarily driven by CommDanger, while CommSafety had a much smaller influence on search food activity. Second, communication displayed opposite effects in the two groups of runs. Otherwise stated, in the presence of predation, animals occasionally take risks in nature due to habituation to alarm signals as discussed in the empirical studies by [16], [17] along with [18]. Whereas in other cases, alarm signals are found to reduce foraging behavior.

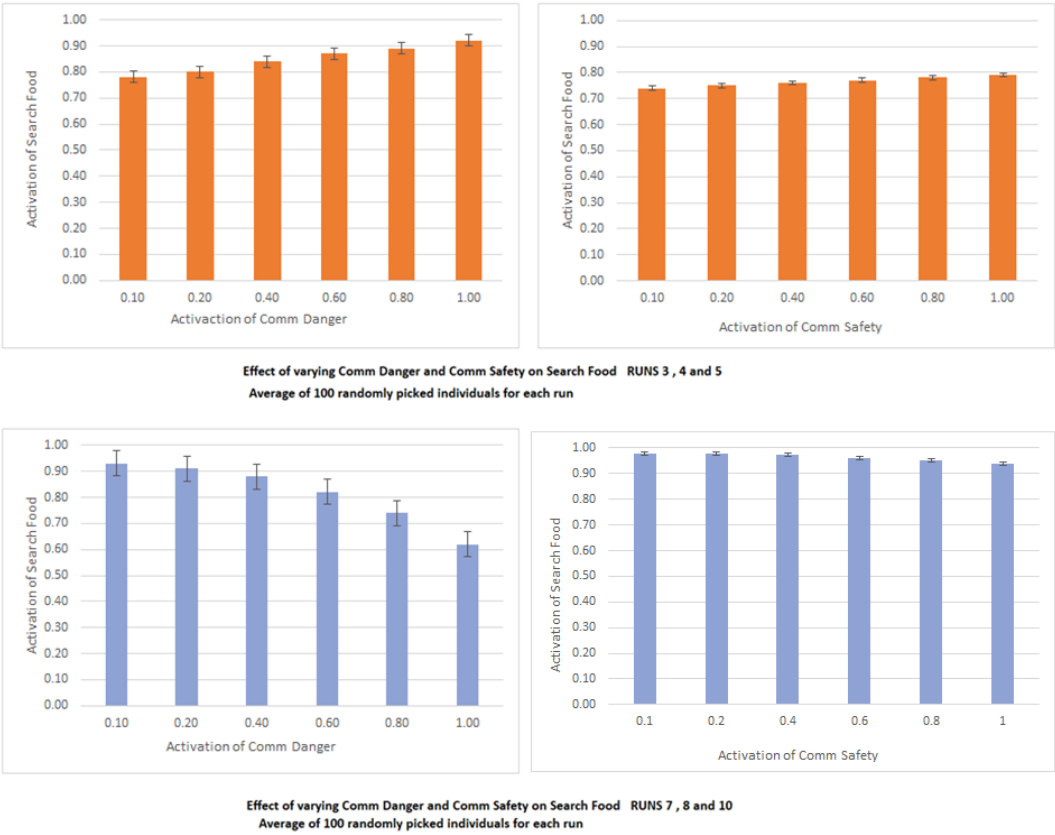


Figure 5.10: FCM Analyzer: Effect of varying CommDanger and CommSafety on SearchFood on individuals from separate runs.

5.3.3 Extraction and analysis of the most significant environmental and behavioral differences between alarm communication and non-communication experiments

We also have attempted to determine the distinguishing factors between Communication and non-communication experiments based on meaningful data related to prey, in addition to specific physical world and individual characteristics. C4.5 decision trees [38] were used to generate the rules based on a subset of features extracted by VAE-EDA-Q [48] as a wrapper method. Raw data collected from each run of EcoSim contained 102 features measuring numerous properties of the world, environment, and the prey individuals at each time step. Using the wrapper method enabled the selection of a subset of only 8 significant features out of 102, thus reducing the complexity and easing the interpretation of the rules that were obtained from the decision trees.

Ten runs for communication and ten runs for non-communication were initially used. For each communication or non-communication run, 30 timesteps were averaged to produce one instance (row) of training data to be used in building the decision trees. The subsequent 20 timesteps were skipped to remove the effects of temporal continuity. To illustrate, the training instance obtained by averaging each of the 102 features from timesteps 30 to 60 would generate the next data point equal to the average of the same features from timesteps 80 to 110 for a given run. Based on the run from which the instance was taken, the training instance was labelled as either communication or non-communication as applicable. Thereafter, all the training instances received from 20 runs using the above averaging procedure were combined into one training dataset, which was used in VAEEDA-Q based Wrapper method.

The effects of communication on foraging behavior have been discussed above. Still, one question that naturally arises scrutinizes the distinguishing features between

Features	Rules Rule code: Communication: C Non-Communication: $\sim$ C AND: $\wedge$
nbPerCell	**Comm-R1:** (nbPerCell $<$ 2.18) $\rightarrow$C
nbPerCell and Concept-Fear	**Comm-R2:** $(ConceptFear > 0.14)$ $\wedge(2.18 \leq nbPerCell < 2.64) \rightarrow C$ **NoComm-R2**: $(ConceptFear > 0.14)$ $\wedge(nbPerCell \geq 2.64) \rightarrow \sim C$
nbPerCell, Concept-Fear and Concept-SearchPartner	**Comm-R3:** $(nbPerCell \geq 2.18)\wedge$ $(Concept - Fear < 0.14)\wedge$ $(Concept - SearchPartner \geq 0.86) \rightarrow C$ **NoComm-R3:** $(nbPerCell \geq 2.18)\wedge$ $(Concept - Fear < 0.14)\wedge$ $(Concept - SearchPartner < 0.86) \rightarrow \sim C$

Table 5.1: Rules obtained from VAE-EDA-Q C4.5 Wrapper

alarm communication runs and non-communication runs. Using machine learning, construction of decision trees elucidated rules that specified the conditions related to alarm communication. Below is a table that outlines the rules derived from the decision trees (see table 5.1).

RULE 1: Throughout the simulation of EcoSim various key parameters related to prey, predators, and the world were recorded to get a meaningful understanding of the unique nature of the run. One such parameter was nbPerCell, which is the average number of individuals per cell and specifies at least one individual measured per timestep. Therefore, it is a direct measure of the density of population. In Communication runs, the individuals are less densely populated. The average number

of prey in each cell is less than 2.18, implying that individuals are more spread in communication runs and more clustered in non -communication runs.

These results cohere with a study by [57] where it was reported that alarm communication decreases as population density increases (dilution of risk) for conspecific collections of zebras. Conversely, [2] found alarm signalling in grey squirrels to be more likely when conspecifics were present with high density than in cases of absence or low density. Urban grey squirrels demonstrated a higher likelihood to use alarm signals when compared to rural grey squirrels, where there is lower density of conspecifics [2]. Authors in [58], however, reported alarm communication in grey squirrels to be completely unaffected by population density. Rule 1 stating that alarm communication favors a population density that is below a certain threshold is outlined in the above table and appears to vindicate the empirical findings of [57]. This aligns with biological philosophy as a high density of individuals is conducive to lessened levels of urgent communication because there are many conspecifics that can offer protection against predators. Contrarily, in less densely populated areas, individuals are on their own, and communication can be indispensable in preserving life in the presence of predators.

RULE 2: Similarly, rule 2 states that communication runs favor population densities with an intermediate range between 2.18 and 2.64 individuals when fear of predation is high (with respect to a given threshold). Non-communication runs favor population densities greater than or equal to 2.64 individuals, when fear of predation is high (with respect to a given threshold). These results are also in agreement with the findings of [57].

RULE 3: Communication runs favor searching for partners when fear of predation is low (with respect to a threshold) and when population density is high (with respect to a threshold). Non-communication runs favor abstaining from the search

for partners under these conditions. This is confirmed by an empirical study conducted by [59] where the authors found the African cichlid fish, Astatotilapia burtoni to engage in acoustic communication during sexual reproduction.

5.3.4 Investigating the possible connection between alarm communication and fitness

Fitness of individuals in EcoSim is measured as the sum of the number of children and grandchildren produced by an individual. This characterization of fitness was developed in [33] and later in [19]. Average fitness of the population was compared in communication runs where communication positively affected search food (foraging), runs where communication negatively affected foraging, and runs where the communication faculty was blocked. Four runs from each test group were considered. The population fitness (average of individuals' inclusive fitness) was measured for each run for 25 timesteps at intervals of 20 timesteps; data in figure 11 represents the average fitness of the population measured over a 500 timestep period taken from the end of each run, consisting of an average of fitness of 25 timesteps in each run.

As discussed in the introduction, evidence from empirical studies suggests a relationship between alarm communication and alterations of foraging behavior and individual fitness; this includes the studies by [20], [22], as well as [23]. In particular, the results of these studies corroborate hypothesis H4 and challenge H3 as outlined above:

H3: Alarm communication augments individual fitness in animal species when habituation to alarm communication occurs.

H4: Alarm communication augments individual fitness in animal species when this communication reduces foraging activity.

As demonstrated by our results in table 5.2, higher levels of fitness are exhibited

Communication Positively Affect Foraging			Communication Negatively Affect Foraging			Blocked Comm		
Run	Avg Fitness	Std. Dev.	Run	Aveg Fitness	Std. Dev.	Run	Avg Fitness	Std. Dev.
1	6.2	8	6	7.3	9.5	1	4.5	10
3	6	9	7	7.1	9	2	4.6	11.5
4	6.1	9.2	8	7.2	9	3	4.3	10
5	6.3	8.5	10	7	9.1	4	4.7	10.6

Table 5.2: Comparing Fitness between runs where Communication positively affects foraging, communication negatively affects foraging and runs where the communications are blocked.

across all runs in connection with communication of both positive and negative influence on foraging as opposed to no communication at all. This appears to corroborate both H3 and H4. However, the results also demonstrate fitness to be higher when alarm communication results in decreased foraging behavior, as opposed to increased foraging behavior, which aligns with H1 and H4 while disputing H3 as outlined above.

5.4 Conclusion

In this study, we used EcoSim, an individual based modeling platform that simulates an artificial ecosystem with three trophic levels (grass, prey, and predators) to model the communication of fear of predation within the prey population. In our simulations we have attempted to observe the effects of the communication of fear on foraging behavior as well as the effects that modifying foraging behavior via communication has on fitness. Further, using machine learning, we attempted to determine the roles that communication plays on population density and reproductive impetus of prey individuals when fear of predators is higher than or lower than a given threshold. To this end, we derived prediction rules from decision trees generated by VAE-EDA-Q C4.5 Wrapper. The hypotheses explored in this work are as follows:

H1: Communication of fear related to predation is associated with a lessening of foraging behavior relative to the absence of communication.

H2: (Predation risk allocation hypothesis) Communication of fear related to predation is associated with an increased level of foraging behavior over time relative to the absence of communication.

H3: Alarm communication augments individual fitness in animal species when there is habituation to alarm communication.

H4: Alarm communication augments individual fitness in animal species when this communication reduces foraging activity.

Based on our simulation results, it appears some cases favor alarm communication that results in increased foraging behavior (possibly due to habituation) thus corroborating H2 outlined above, while other cases favor alarm communication that results in decreased foraging behavior, thereby corroborating H1 as outlined above. As outlined in sections 5.3.1 and 5.3.2, empirical evidence exists in support of both effects of alarm communication. Although one may expect an increase in inclusive fitness for both cases, it would appear from our data that higher fitness is a product of cases where alarm communication results in decreased foraging as opposed to cases where alarm communication results in increased foraging. This data corroborates H4 and challenges H3. In addition to the number of empirical studies outlined in section 5.3.4 that validate our findings regarding fitness. Biological theory also supports our findings since foraging in the presence of predation may have a negative effect on survival if the predators are largely successful. Under this philosophy, individuals would be unable to reproduce and pass on their genes to future generations. Using machine learning, we were able to determine the features that characterize alarm communication runs. First, we discovered that communication runs favor a low population density relative to a given threshold, which has been validated by a number of em-

pirical studies. Notwithstanding, we also discovered that communication runs favor population densities intermediate of two thresholds t1 and t2 (where t1 $<$ t2) when fear is higher than a given threshold (0.14). Non-communication, however, were found to favor population densities that exceed the higher threshold t2 when fear is higher than the threshold value 0.14.

Alarm communication appears to occur in lower population densities when there a higher level of fear of predation occurs, potentially due to the presence of fewer individuals to warn others. Higher density populations, according to [57], exercise a so-called dilution of risk. Finally, we discovered that communication runs favor the desire to search for reproductive partners when fear is below a given threshold and population density is higher than a given threshold. Conversely, non-communication runs do not favor searching for reproductive partners under these conditions. This is supported by biological theory, since a greater level of conspecifics is conducive to a higher chance of socialization and eventual reproduction, especially when there is reduced fear of predation. Further, socialization for the purpose of reproduction has a greater chance of success when communication between the individuals is involved.

Bibliography

[1] Reginald B Cocroft and Rafael L Rodríguez. The behavioral ecology of insect vibrational communication. *AIBS Bulletin*, 55(4):323–334, 2005.

[2] Sarah R Partan, Andrew G Fulmer, Maya AM Gounard, and Jake E Redmond. Multimodal alarm behavior in urban and rural gray squirrels studied by means of observation and a mechanical robot. *Current Zoology*, 56(3), 2010.

[3] Anna Dornhaus and Lars Chittka. Food alert in bumblebees (Bombus terrestris): possible mechanisms and evolutionary implications. *Behavioral Ecology and So-*

ciobiology, 50(6):570–576, 2001.

[4] Adam J Leavesley and Robert D Magrath. Communicating about danger: urgency alarm calling in a bird. *Animal behaviour*, 70(2):365–373, 2005.

[5] Delbert L Smee and Marc J Weissburg. Hard clams (Mercenaria mercenaria) evaluate predation risk using chemical signals from predators and injured conspecifics. *Journal of chemical ecology*, 32(3):605–619, 2006.

[6] Matthew D Mitchell, Peter F Cowman, and Mark I McCormick. Chemical alarm cues are conserved within the coral reef fish family Pomacentridae. *Plos one*, 7(10):e47428, 2012.

[7] R Inta, TA Evans, and JCS Lai. Effect of vibratory soldier alarm signals on the foraging behavior of subterranean termites (Isoptera: Rhinotermitidae). *Journal of economic entomology*, 102(1):121–126, 2009.

[8] A Randall Hughes, David A Mann, and David L Kimbro. Predatory fish sounds can alter crab foraging behaviour and influence bivalve abundance. *Proc. R. Soc. B*, 281(1788):20140715, 2014.

[9] Bradley F Dickey and Thomas M McCarthy. Predator–prey interactions between crayfish (Orconectes juvenilis) and snails (Physa gyrina) are affected by spatial scale and chemical cues. *Invertebrate Biology*, 126(1):57–66, 2007.

[10] Michael S Pollock, Xiaoxia Zhao, Grant E Brown, Robin C Kusch, Robyn J Pollock, and Douglas P Chivers. The response of convict cichlids to chemical alarm cues: an integrated study of behaviour, growth and reproduction. In *Annales Zoologici Fennici*, pages 485–495. JSTOR, 2005.

[11] Andrew Sih, Robert Ziemba, and Karin C Harding. New insights on how temporal variation in predation risk shapes prey behavior. *Trends in Ecology & Evolution*, 15(1):3–4, 2000.

[12] Maud CO Ferrari, François Messier, and Douglas P Chivers. Variable predation risk and the dynamic nature of mosquito antipredator responses to chemical alarm cues. *Chemoecology*, 17(4):223–229, 2007.

[13] Steven L Lima and Peter A Bednekoff. Temporal variation in danger drives antipredator behavior: the predation risk allocation hypothesis. *The American Naturalist*, 153(6):649–659, 1999.

[14] Guy Beauchamp and Graeme D Ruxton. A reassessment of the predation risk allocation hypothesis: a comment on Lima and Bednekoff. *The American Naturalist*, 177(1):143–146, 2010.

[15] Maud CO Ferrari, Andrew Sih, and Douglas P Chivers. The paradox of risk allocation: a review and prospectus. *Animal Behaviour*, 78(3):579–585, 2009.

[16] Andrew Sih and Thomas M McCarthy. Prey responses to pulses of risk and safety: testing the risk allocation hypothesis. *Animal Behaviour*, 63(3):437–443, 2002.

[17] Reehan S Mirza, Alicia Mathis, and Douglas P Chivers. Does temporal variation in predation risk influence the intensity of antipredator responses? A test of the risk allocation hypothesis. *Ethology*, 112(1):44–51, 2006.

[18] Stefanie Slos and Robby Stoks. Behavioural correlations may cause partial support for the risk allocation hypothesis in damselfly larvae. *Ethology*, 112(2):143–151, 2006.

[19] Brendan P Zietsch, Ralf Kuja-Halkola, Hasse Walum, and Karin JH Verweij. Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. *Proceedings of the National Academy of Sciences*, 111(3):1032–1036, 2014.

[20] Martin De Vos, Wing Yin Cheng, Holly E Summers, Robert A Raguso, and Georg Jander. Alarm pheromone habituation in Myzus persicae has fitness consequences and causes extensive gene expression changes. *Proceedings of the National Academy of Sciences*, 107(33):14673–14678, 2010.

[21] Denis Meuthen, Sebastian A Baldauf, and Timo Thünken. Evolution of alarm cues: a role for kin selection. *F1000 Research*, 1(27):1–27, 2014.

[22] Grant E Brown, Jean-Francois Poirier, and James C Adrian Jr. Assessment of local predation risk: the role of subthreshold concentrations of chemical alarm cues. *Behavioral Ecology*, 15(5):810–815, 2004.

[23] Klaus Zuberbühler, David Jenny, and Redouan Bshary. The predator deterrence function of primate alarm calls. *Ethology*, 105(6):477–490, 1999.

[24] Erin L Shelley and Daniel T Blumstein. The evolution of vocal alarm communication in rodents. *Behavioral Ecology*, 16(1):169–177, 2004.

[25] Daniel T Blumstein. The evolution of alarm communication in rodents: structure, function, and the puzzle of apparently altruistic calling. *Rodent Societies: an Ecological & Evolutionary Perspective*, page 317e327, 2007.

[26] Robin Gras, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life*, 15(4):423–463, 2009.

[27] Morteza Mashayekhi, Brian MacPherson, and Robin Gras. A machine learning approach to investigate the reasons behind species extinction. *Ecological informatics*, 20:58–66, 2014.

[28] Abbas Golestani, Robin Gras, and Melania Cristescu. Speciation with gene flow in a heterogeneous virtual world: can physical obstacles accelerate speciation? *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1740):3055–3064, 2012.

[29] Marwa Khater, Dorian Murariu, and Robin Gras. Contemporary evolution and genetic change of prey as a response to predator removal. *Ecological informatics*, 22:13–22, 2014.

[30] Maryam Karim Pour, Sourodeep Bhattacharjee, Robin Gras, and Ken Drouillard. The integration of an individual-based model into toxicokinetics to enhance ecological realism in evaluating population-level impacts of exposure to PCB. In *Complex Systems (WCCS), 2015 Third World Conference on*, pages 1–8. IEEE, 2015.

[31] Sourodeep Bhattacharjee, Brian MacPherson, and Robin Gras. A comparison of sexual selection versus random selection with respect to extinction and speciation rates using individual based modeling and machine learning. *Ecological Complexity*, 36:126–137, 2018.

[32] Bart Kosko. Fuzzy cognitive maps. *International journal of man-machine studies*, 24(1):65–75, 1986.

[33] Jon E Brommer, Lars Gustafsson, Hannu Pietiäinen, and Juha Merilä. Single-generation estimates of individual fitness as proxies for long-term genetic contribution. *The American Naturalist*, 163(4):505–517, 2004.

[34] Brian MacPherson, Morteza Mashayekhi, Robin Gras, and Ryan Scott. Exploring the connection between emergent animal personality and fitness using a novel individual-based model and decision tree approach. *Ecological Informatics*, 40:81–92, 2017.

[35] Lisa Teckentrup, Volker Grimm, Stephanie Kramer-Schadt, and Florian Jeltsch. Community consequences of foraging under fear. *Ecological Modelling*, 383:80–90, 2018.

[36] Sreerama K Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4):345–389, 1998.

[37] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[38] J Ross Quinlan. C4. 5: Programming for machine learning. *Morgan Kauffmann*, 38:48, 1993.

[39] Mark A Hall and Lloyd A Smith. Practical feature subset selection for machine learning. In *Computer science'98 proceedings of the 21st Australasian computer science conference ACSC*, volume 98, pages 181–191, 1998.

[40] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[41] Shumeet Baluja. Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science, 1994.

[42] Heinz Mühlenbein and Gerhard Paass. From recombination of genes to the estimation of distributions I. Binary parameters. In *International conference on parallel problem solving from nature*, pages 178–187. Springer, 1996.

[43] Pedro Larrañaga and Jose A Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.

[44] Martin Pelikan, David E Goldberg, and Fernando G Lobo. A survey of optimization by building and using probabilistic models. *Computational optimization and applications*, 21(1):5–20, 2002.

[45] Iñaki Inza, Pedro Larrañaga, Ramón Etxeberria, and Basilio Sierra. Feature subset selection by Bayesian network-based optimization. *Artificial intelligence*, 123(1-2):157–184, 2000.

[46] Marwa Khater, Elham Salehi, and Robin Gras. Correlation between genetic diversity and fitness in a predator-prey ecosystem simulation. In *Australasian Joint Conference on Artificial Intelligence*, pages 422–431. Springer, 2011.

[47] Elham Salehi and Robin Gras. Efficient EDA for large opimization problems via constraining the search space of models. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 73–74. ACM, 2011.

[48] Sourodeep Bhattacharjee and Robin Gras. Estimation of Distribution using Population Queue based Variational Autoencoders. *In Press to IEEE CONGRESS ON EVOLUTIONARY COMPUTATION*, 2019.

[49] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[50] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic back-propagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[51] Luis-Miguel Chevin, Sinéad Collins, and François Lefèvre. Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field. *Functional Ecology*, 27(4):967–979, 2013.

[52] Theodore Garland and Scott A Kelly. Phenotypic plasticity and experimental evolution. *Journal of Experimental Biology*, 209(12):2344–2361, 2006.

[53] Trevor D Price, Anna Qvarnström, and Darren E Irwin. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1523):1433–1440, 2003.

[54] Andrew P Hendry. *Eco-evolutionary dynamics*. Princeton university press, 2016.

[55] David A Reznick, Heather Bryga, and John A Endler. Experimentally induced life-history evolution in a natural population. *Nature*, 346(6282):357, 1990.

[56] R. Scott, B. MacPherson, and R. EcoSim Gras. EcoSim, an enhanced artificial ecosystem: addressing deeper behavioral, ecological, and evolutionary questions. *Cognitive Architectures - In Press*, 2018.

[57] Melissa H Schmitt, Keenan Stears, Christopher C Wilmers, and Adrian M Shrader. Determining the relative importance of dilution and detection for zebra foraging in mixed-species herds. *Animal behaviour*, 96:151–158, 2014.

[58] Ilya A Volodin, Elena V Volodina, Vera A Matrosova, Lyudmila E Savinetskaya, Olga N Shekarova, and Vera A Voytsik. Population density does not affect the

alarm call characteristics in the speckled ground squirrel Spermophilus suslicus. *Lynx (Praha), ns*, 39(2):333–342, 2008.

[59] Karen P Maruska, Uyhun S Ung, and Russell D Fernald. The African cichlid fish Astatotilapia burtoni uses acoustic communication for reproduction: sound production, hearing, and behavioral significance. *PLoS One*, 7(5):e37612, 2012.

Chapter 6

Conclusions

The investigation of natural systems within the real world has historically presented challenges for scientists as complexities involved with ecological phenomena are fascinating, yet difficult to understand. Artificial simulations have recently unburdened the study of natural systems through modeling approaches that elucidate the behavior of complex ecosystems. Individual based modeling creates virtual ecosystems wherein a bottom-up approach allows traits and behaviors of individual organisms to be held at the forefront and summate to overall outcomes based on interactions. Due to the very nature of this approach, myriads of interactions have the propensity to produce non-linear behaviors that generate vast amounts of data. Thus, with the objective of transforming raw data into insight, machine learning is employed to extract useful knowledge, suggest conclusions, and aid in the process of decision making through learning from the input raw data. In tandem, individual based modeling and machine learning make possible the understanding of high-level, true to nature, ecological phenomena [1].

Several machine learning methods exist to mitigate issues involved with high dimensional datasets. One approach, feature selection, focuses on the removal of ex-

traneous features prior to processing and results in concise models produced with a reasonable computational resource budget. One approach to feature selection, called wrapper methods, integrates the feature subset and classifier model search; this leads to the generation of undeniably superior classification models. Estimation of distribution algorithms, in the capacity of efficient continuous and non-linear optimizations, have demonstrated success for use as a wrapper method for feature subset search and selection.

We constructed a novel estimation of distribution algorithms (EDA) based upon the combination of two novel generative models known as variational autoencoder model building algorithms. Variational Autoencoder with Population Queue (VAE-EDA-Q) relies upon a queue of historical populations, updated at each iteration of EDA, to smooth the data generation process. Adaptive Variance Scaling (AVS) dynamically updates the variance at which probabilistic models are sampled based on the fitness of solutions. Cooperatively, the effects of our method of combining VAE-EDA-Q with AVS, demonstrated high computational efficiency and required few fitness evaluations when we tested them on Trap-k and NK landscapes [2]. In addition to benchmark problems, we tested VAE-EDA-Q AVS on CIFAR10 and CIFAR100 datasets for a real world comparison to state-of-the-art CNN algorithms, of which were either fully automated, manually designed, or required partial human intervention. VAE-EDA-Q AVS demonstrated an average of 1.5% higher accuracy on CIFAR10 than all other state of the art algorithms while requiring 25% less parameters and an average of 6% higher accuracy on CIFAR100 than all other state of the art algorithms using 10% less parameters. Further, VAE-EDA-Q demonstrated its proficiency in exploring continuous latent space representation of features, which enables randomized sampling and smooth interpolation over the manifold of high-quality solutions within the feature space. Accordingly, we used VAE-EDA-Q as a wrapper

method to reduce the number of features on results obtained from EcoSim [3].

Through the optimization of machine learning approaches employed within EcoSim, several highly debated concepts were able to be resolved by modeling and examining intricate relationships. The relationship between selective mating, speciation rate, and extinction rate was one concept explicated through individual based computer simulations that had previously lacked consonance based on empirical study. Simulations produced by EcoSim demonstrated a significant difference between sexual selection and random selection as related to speciation rates, extinction rates, and species turnover rates. Machine learning was employed to generate prediction rules regarding rates of speciation and extinction for both sexual and random selection groups. Within random selection, speciation was found to be significantly higher.

Animal alarm communication, as related to foraging behavior and fitness, was another topic to be resolved through individual based modeling. Results demonstrated alarm communication to decrease foraging activity in most cases, yet gradually increase foraging activity in the case of habituation. Furthermore, both outcomes resulting from alarm communication were found to increase fitness as compared to non-communication. Additionally, within alarm communication, fitness was found to be higher within groups that displayed decreased foraging activity as compared to groups that gradually increased foraging activity in response to habituation. Lastly, machine learning was employed to discern features categorizing communicative or non-communicative environments, from which conclusions could be formed about the impact of communication on several factors related to a community of prey. Communication was found to be important in cases of low population density and high reproductive urgency in the context of low levels of fear. Communication was found to be less important in the presence of high numbers of conspecifics. This study highlights the interest of machine learning and individual based modeling as specific,

intricate rules and resolutions were able to be determined through the testing of four hypotheses.

As discovery progresses within the field of machine learning, the study of natural systems is expected to advance to new territory. Future work proposed for the advancement of VAE-EDA-Q AVS involves dynamically altering the population queue size in accordance with the state of iteration, as well as balancing the amounts of exploration and exploitation as the algorithm progresses [2]. Future extensions related to sexual selection involve the examination of long-term evolutionary consequences. Using a machine learning approach similar to that employed by authors in [4], distinct, specific evolutionary patterns relative to sexual selection and random mating could be studied by altering the sexual selection policy. Long runs could be executed to observe and compare short and long term patterns. Further, the same approach could be applied to real ecological and paleontological data to generate rules for the analysis of ecological and evolutionary effects of sexual selection in this capacity.

As related to the study of communication, future extensions propose to explore the possibility of communication of other internal concepts such as curiosity and satisfaction. Using the approach outlined in chapter 5, the relationship between these concepts and other individuals and the environment could be elucidated. Further, specific internal concepts such as hunger and chasing prey could be studied as effectors of hunting behavior and fitness of predators. Advanced applications could examine communication of entire subsets of internal concepts to ascertain effectual concepts for survival and fitness.

These studies, and many similar, have harnessed the capability of individual based modeling in conjunction with machine learning to reach outcomes that previously required a lifetime to realize. The validation of the results with real-world observations has confirmed the significance of their use.

Bibliography

[1] Morteza Mashayekhi. Individual-Based Modeling and Data Analysis of Ecological Systems Using Machine Learning Techniques. 2015.

[2] Sourodeep Bhattacharjee and Robin Gras. Estimation of Distribution using Population Queue based Variational Autoencoders. *In Press, IEEE CONGRESS ON EVOLUTIONARY COMPUTATION*, 2019.

[3] Robin Gras, Didier Devaurs, Adrianna Wozniak, and Adam Aspinall. An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial life*, 15(4):423–463, 2009.

[4] Sourodeep Bhattacharjee, Brian MacPherson, and Robin Gras. A comparison of sexual selection versus random selection with respect to extinction and speciation rates using individual based modeling and machine learning. *Ecological Complexity*, 36:126–137, 2018.

Appendix A

Edges of FCM

Range of Edge weight

1						
2	Desc	-1 to -0.6	-0.5 to -0.1	0	0.1 to 0.5	0.6 to 1
3	CommDanger -> Fear	42	1740	2966309	1242	25
4	CommDanger -> Hunger	92	2048	2964071		
5	CommDanger -> SearchPartner	123	1414	2965695		
6	CommDanger -> Curiosity	120	4000	2963585		
7	CommDanger -> Sedentary	68	2092	2965543		
8	CommDanger -> Satisfaction	23	1370	2963917		
9	CommDanger -> Nuisance	53	1996	2964739		

Number of Individuals using the given edge (here CommDanger->Satisfaction) with given weight range

(Here -1 to -0.6)

Step 1: We begin with the above data which relates to a particular timestep of a particular run.

Step 2: We find a combined index of popularity/importance of a given edge as follows:

Index = ∑ (Edge Weight * Number of Individuals)

The higher limit of an edge weight range is multiplied by the number of individuals using that edge. Thereafter we add the values for a given edge

In the image below for the first row (CommDanger -> Fear) we find index as follows:

Index = (-1 * 42) + (-0.5 * 1740) +(0.5 * 1242) + (1 * 25)

= 884.5

This index is denoted as C_D for CommDanger connections and C_S for CommSafety connections

						The value of this cell is given by Sum of product of resp. Edge Weight and Population (CI_D for Comm Danger and CI_S for Comm Safety): (-1 * B2) + (-0.5 * C2) + (0.5 * E2) + (1*F2)
Desc	-1 to -0.6	-0.5 to -0.1	0	0.1 to 0.5	0.6 to 1	Edge Index
CommDanger -> Fear	42	1740	2966309	1242	25	884.5
CommDanger -> Hunger	92	2048	2964071	3111	36	475.5
CommDanger -> SearchPartner	123	1414	2965695	2096	30	248
CommDanger -> Curiosity	120	4000	2963585	1623	30	-1278.5
CommDanger -> Sedentary	68	2092	2965543	1608	47	-263
CommDanger -> Satisfaction	23	1370	2963917	4012	36	1334
CommDanger -> Nuisance	53	1996	2964739	1684	886	677
CommDanger -> Escape	34	1931	2963512	3715	166	1024
CommDanger -> SearchFood	86	2734	2965021	1497	20	-684.5
CommDanger -> Socialize	102	1537	2965446	2233	40	286
CommDanger -> Exploration	71	4412	2963552	1297	26	-1602.5

Step 3:

The combined index is calculated as C_D - C_S

The value of this cell is given by Sum of product of resp. Edge Weight and Population (CI_D for Comm Danger and CI_S for Comm Safety): (-1 * B2) + (-0.5 * C2) + (0.5 * E2) + (1*F2)		(CI_D - CI_S)
Edge Index	CommSafety Values	**CommDanger-CommSafety Combined Index**
884.5	-125.5	**1010**
475.5	-1592.5	**2068**
248	831.5	**-583.5**
-1278.5	-738.5	**-540**
-263	1124	**-1387**
1334	-224	**1558**
677	557	**120**
1024	-213.5	**1237.5**
-684.5	-1107	**422.5**
286	85	**201**
-1602.5	-218	**-1384.5**
746.5	629	**117.5**
-138.5	-544	**405.5**
-1660	-1060.5	**-599.5**
377	336	**41**
-80.5	220	**-300.5**
-241	-256	**15**

Vita Auctoris

Sourodeep Bhattacharjee was born in 1987 in Howrah, India. He received his Bachelor of Technology and Master of Science Degrees in Computer Science and Engineering from West Bengal University of Technology, India and University of Windsor, Canada, in 2010 and 2012 respectively. He studied Doctor of Philosophy in the School of Computer Science, University of Windsor, Canada, from 2014 to 2019. His research interests include Machine Learning, Combinatorial Optimizations and Ecological Modeling.