Optimizing Ecological Sustainability by Integrating Intuition and Machine Learning via Gamification

Ben Saylor[1], Anagha Kulkarni[1], Neo Martinez[2], Ilmi Yoon[1]
bsaylor@mail.sfsu.edu, ak@sfsu.edu, neomartinez@email.arizona.edu, ilmi@sfsu.edu
[1]San Francisco State University, [2]University of Arizona

1. Introduction

Improving human-computer interactions can help increase critical understanding of ecological sustainability [1] by, for example, helping to parameterize dynamical systems models of the complex structure and nonlinear dynamics of ecological systems. A prominent family of such models includes allometric trophic network (ATN) models [2,3], which use metabolic rates estimated from body size and the networks of feeding relationships among species to describe the flow of energy and biomass through food webs. An ATN model calculates species' changes in biomass as gains due to photosynthesis and feeding minus losses due to metabolism and being consumed. ATN models consist of systems of differential equations with several parameters for each species such as metabolic rate and also for each link such as assimilation efficiency. Given the large parameter space of the many variables and their interactions, parameterizing ATN models so that they (1) dynamically sustain many species and (2) accurately predict organismal abundance in nature presents significant challenges. We address these challenges by gamifying ATN models so that human intuition and visual reasoning are better integrated with machine learning and computational power in order to better understand ecological sustainability.

2. Convergence Game

Our Convergence game is one of several "World of Balance" [4] (WoB) mini games where players act as stewards of virtual ecosystems based on the well-studied Serengeti food web. Our original Convergence game challenges players to closely match a time series of different species' biomass in a simulated food web by parameterizing an ATN model. Players are shown a default graph and a target graph of data produced by parameterizing an ATN model with two different sets of values. Players manipulate a set of sliders to change the default parameter values in order to make the default graph look more like the target graph and win credits. The parameters that can be manipulated in the game are the metabolic rate *X* for animal species and carrying capacity *K* for plant species.

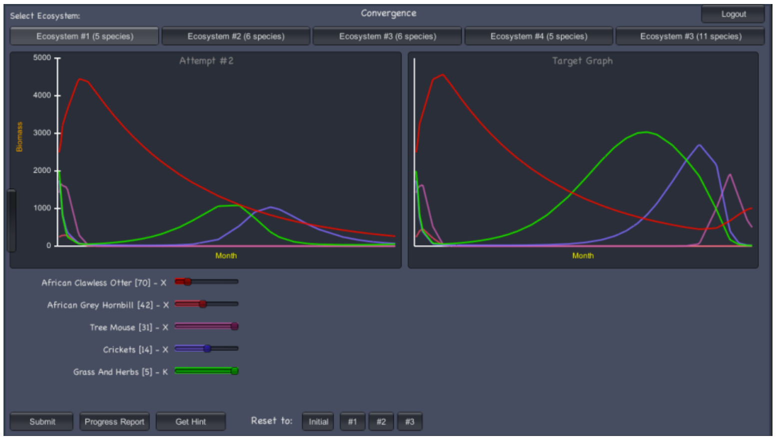

Figure 1: Convergence game screen

Our current Multiplayer Convergence game challenges two players to compete for the best match between the default and target graphs. Gameplay takes place across several rounds in which players bet with game credits that their parameter manipulations match better than the other player's manipulations. The winner of each round receives half of the pot. The other half is accumulated until the end of the game, which is won by the winner of the final round.

3. Problem definition

Even in a small food web consisting of a handful of species the parameter space defined by the minimal set of parameters is fairly large. Due to the network of interdependencies within a food web, a small change made to one species' parameter impacts the dynamics of all species, and thus the ecosystem. This makes it non-trivial to find the parameter values that match the default to the target graph. This difficulty motivates our central goal of developing an intuitive understanding of ecological dynamics and sustainability by supporting players to discover parameter values that replicate the target ecosystem.

4. Methodology

To produce the target graphs for Multiplayer Convergence, we sought to improve the data used for the original Convergence game and to determine ranges of parameter values likely to produce desired outcomes and be displayed to players as hints. We achieved this using a 4-step process for each target graph: (1) species selection, (2) parameter space exploration and simulation, (3) machine-learning classification of simulation results, and (4) evaluation of derived parameter ranges.

4.1 Species selection: We represent food webs as directed graphs in which nodes represent species and an edge (*u*, *v*) indicates that species *u* is eaten by species *v*. We maintained manageable levels of complexity by constructing smaller food webs based on the full WoB food web with 87 species using a sampling algorithm that produces subgraphs with the following properties: (1) having exactly one connected component, (2) containing all possible trophic links from the full web involving the species in the sub-web, (3) having a specified number of plant species deemed appropriate for the sub-web size, and (4) having no incomplete food chains – that is, having no non-plants as source nodes. Additionally, the algorithm minimizes the number of plant-eating species lacking predators in the sub-web.

4.2 Parameter space exploration and simulation: For each sub-web, we ran 1000 simulations using the ATN Engine with randomized parameters. We independently drew each parameter value from a uniform distribution between 50% and 150% of its default value. The WoB database contains default parameter values for each species based on measurements and established models of metabolic rates.

4.3 Machine-learning classification of simulation results: We used the simulation data to train a machine learning algorithm to predict ecosystems health based on model parameters. For each simulation, we calculated the linear time trend of the "environment score." Since predators at high trophic levels such as raptors and lions indicate ecosystem health, the score is the total biomass of the ecosystem, biased such that biomass at high trophic levels counts more than biomass at low levels. Score trend values were assigned as "good" or "bad" (or no label) if they were among the top 25% or bottom 25% (or middle 50%) of simulations. The simulation's input parameters and this ecosystem health label comprised training data for decision tree classification using Weka's implementation of the C4.5 algorithm [5,6]. While this provided decision trees capable of predicting ecosystem health with a high degree of accuracy, we were more interested in obtaining information about promising parameter ranges making predictions for specific combinations of parameter values.

```
X31 <= 0.7
|   X42 <= 0.637
|   |   X70 <= 0.1395: bad (41.0)
|   |   X70 > 0.1395: good (87.0)
|   X42 > 0.637: bad (138.0)
X31 > 0.7
|   X70 <= 0.217: bad (3866.0/3.0)
|   X70 > 0.217
|   |   R5 <= 0.8: good (11.0)
|   |   R5 > 0.8: bad (19.0)
```

Figure 2. A decision tree

4.4 Evaluation of parameter ranges: We used the structure of the decision trees to derive and evaluate parameter ranges. Each node in a binary decision tree represents a branching criterion on the input attributes that distinguishes the data instances in the left subtree from those in the right subtree. These criteria are chosen by the decision tree learning algorithm to maximize the predictive accuracy of the tree on the training data. For real-valued input attributes such as the ATN parameters, each criterion consists of parameter name and a threshold value that separates the instances in the two subtrees. Thus, for each each parameter there is a (possibly empty) set of values, given by the decision tree nodes which have been chosen to effectively separate "good" from "bad" instances. We derived ranges for each parameter from each pair of successive threshold values listed in sorted order. Each range was assigned a score equal to the probability that a simulation drawn from within the range has a label of "good," minus the probability that it has a "bad" label. Ranges with a score above a certain threshold are shown to the players as hints.

5. Experimental Design

In our user study, planned for the fall of 2016, we will compare two separate groups of Multiplayer Convergence players. One group plays with hints (parameter ranges likely to lead to healthy ecosystems) and the other group plays without such hints. This design will enable us to experimentally study human-computer interactions with a focus on the effect of hints derived from the environment scores on how quickly and accurately players discover parameter sets that generate biomass dynamics match the dynamics described in the target graph.

6. Conclusion

We have established a foundation upon which we can test various hypotheses about guiding Convergence players using information about the parameter search space derived from machine learning. As we continue to develop the tools, methodology, and experimental design, we will explore further possibilities of integrating machine learning and gamification in order to parameterize dynamic models of complex ecological networks and better understand the sustainability of the ecosystems the models represent.

References

[1] Martinez, N.D., P. Tonin, B. Bauer, R.C. Rael, S. Yoon, I. Yoon, J.A. Dunne. 2012. Sustaining economic exploitation of complex ecosystems in computational models of coupled human-natural networks. *AAAI* 2012. *Proceedings of The Association for the Advancement of Artificial Intelligence*.

[2] Alice Boit, Neo D Martinez, Richard J Williams, and Ursula Gaedke, *Mechanistic theory and modelling of complex food-web dynamics in Lake Constance*, Ecology letters **15** (2012), no. 6, 594–602.

[3] Kuparinen, A., F. S. Valdovinos, A. Boit, H. Lassaux and Neo D. Martinez. 2016. Fishing-induced life-history changes degrade and destabilize fishery ecosystems. *Scientific Reports* 6, 22245.

[4] Ilmi Yoon, Gary Ng, Hunvil Rodrigues, Trang Nguyen, Jae H Paik, Sanghyuk Yoon, Rich Williams, and Neo D Martinez, *Iterative design and development of the 'World of Balance' game: from ecosystem education to scientific discovery*, Games Innovation Conference (IGIC), 2013 IEEE International, IEEE, 2013, pp. 283–290.

[5] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[6] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.