Unsupervised Machine Learning in Agent-Based Modeling

Luke Robinson and Forrest Stonedahl*

Mathematics and Computer Science Department, Augustana College   *Faculty Advisor

I. Overview

Agent-based models (ABMs) are used by researchers in a variety of fields to model natural phenomena. In an ABM, a wide range of behaviors and outcomes can be observed based on the parameters of the model. In many cases, these behaviors can be categorized into discrete outcomes identifiable by human observers. Our goal was to use clustering algorithms to identify those outcomes from model output data. If this task can be completed reliably by a computer, it will make the task of investigating an ABM easier for human users. For this project, we used the Wolf Sheep Predation model that can be found in the models library of Netlogo 5.3.1 [2]. For data analysis and visualization, we used Python 3.5.2 and the scikit-learn and matplotlib libraries.

II. Agent-Based Model

Agent-based modeling is "a form of computational modeling whereby a phenomenon is modeled in terms of agents and their interactions" [1]. In the Wolf Sheep Predation model, wolves (black) and sheep (white) move randomly on a field of grass and dirt patches. When a wolf and a sheep occupy the same space, the wolf eats the sheep. If a wolf or sheep consumes a sufficient amount of food, it reproduces. If a wolf or sheep goes long enough without food, it dies out. A final detail: If the 'grass?' toggle is ON, grass can be depleted by the sheep (resulting in a brown patch, which will eventually grow back). If 'grass?' is OFF, then grass is an infinite resource for the sheep.
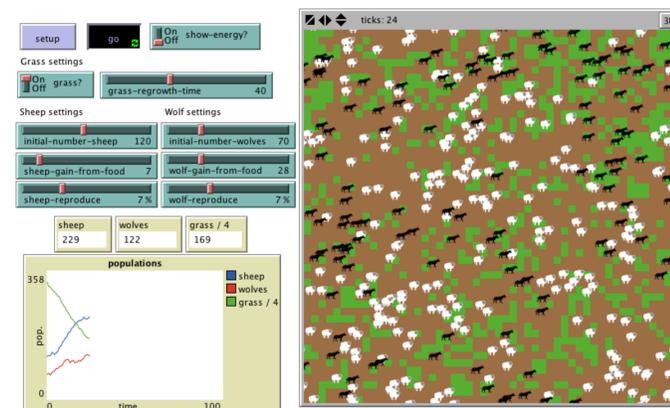
BehaviorSpace Experiment

BehaviorSpace is a tool in Netlogo that allows the user to perform an experiment that carries out many runs of a model with different combinations of the model's parameters. The data for this project came from a BehaviorSpace experiment with the following parameters:

["sheep-reproduce" 3 5 7]
["initial-number-sheep" [80 10 120]]
["wolf-gain-from-food" [12 8 28]]
["wolf-reproduce" 3 5 7]
["show-energy?" false]
["sheep-gain-from-food" 3 5 7]
["initial-number-wolves" [30 10 70]]
["grass-regrowth-time" [20 10 40]]
["grass?" true false]

Output measures:
count sheep,
count wolves
Stop condition:
not any? turtles
or count sheep > 1500
Time limit: 1500 steps
Total runs: 24,300

III. Clustering Algorithms

K-Means($n\_clusters$=2)

silhouette score = .86

DBSCAN($epsilon$=50, $min\_samples$=30)

silhouette score = .66

AffinityPropagation($damping$=.75)

silhouette score = .60

K-Means

K-Means requires the user to specify the number of clusters, and it clusters data by trying to separate samples into groups of equal variance [4]. Our data forms two clusters: simulations where the sheep population exploded, and simulations where it didn't. K-Means generally found these two clusters, but misclassified some points on the boundary (the blue points between 800 and 1100 sheep).

DBSCAN

DBSCAN is short for 'Density-Based Spatial Clustering of Applications with Noise' [3]. This algorithm finds core samples of high density and expands clusters from these. An $epsilon$ value of 50 sets the maximum distance between two points to be in the same neighborhood. A value of 30 for $min\_samples$ specifies how many data points must be in a neighborhood for it to be considered the core of a cluster. The black dots on the diagonal fringe were classified as noise (not part of any cluster).

Affinity Propagation

The Affinity Propagation clustering algorithm chooses 'exemplars' (shown as large dots at left), which are points that best represent others around them, and then forms clusters by growing them outwards from the exemplars [4]. Due to the time complexity of this algorithm and the large dataset that we used, we had to run the algorithm on a random subset with only 2,000 data points. The sparser data set contributed to why this algorithm found more clusters than the other algorithms.

Silhouette Scores

The silhouette score is a clustering metric ranging from -1 (worst) to 1 (best), by comparing distances between points within a cluster and between different clusters [4].
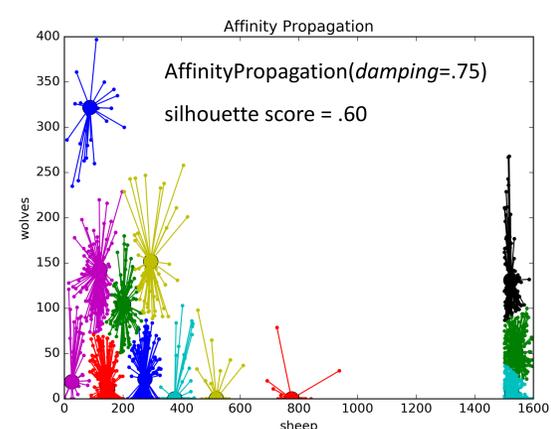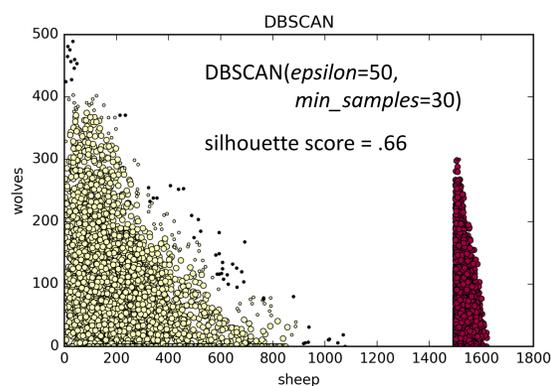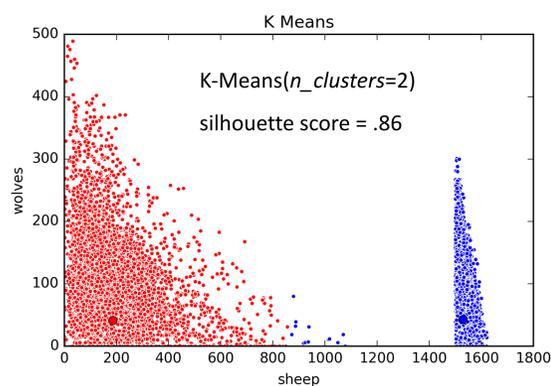
IV. Parameter Analysis

We investigated the two clusters that made up the K-means clustering by examining the distribution of the parameters from the ABM that led to a run being in its respective cluster.

The histograms for most of the variables showed an even distribution for both clusters, with the notable exceptions of the 'wolfgainfromfood' and 'grass' variables.
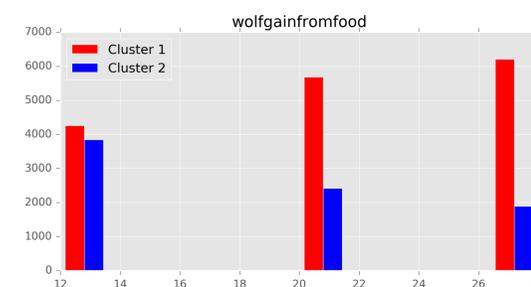

When 'wolfgainfromfood' has a higher value, the run is more likely to be in cluster 1 (red), where the sheep population did not explode. The reverse is also true. This makes sense because presumably the wolves' increased gain from food enabled them to grow in population and limit the sheep population growth.
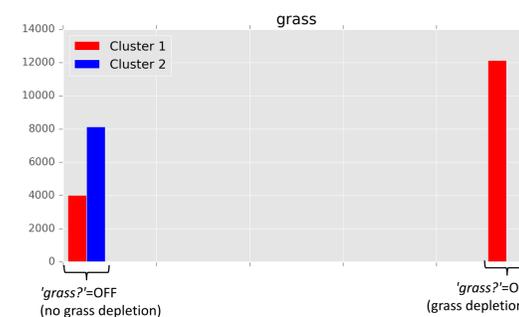
In cluster 1, ¾ of the trials had grass-depletion enabled.
In cluster 2, basically all trials had grass depletion disabled.
This points to a key outcome from this model, which is that an ecosystem with three interacting species is more stable (less likely to have a population explosion) than an 2-species ecosystem.

V. Overarching Goals and Future Work
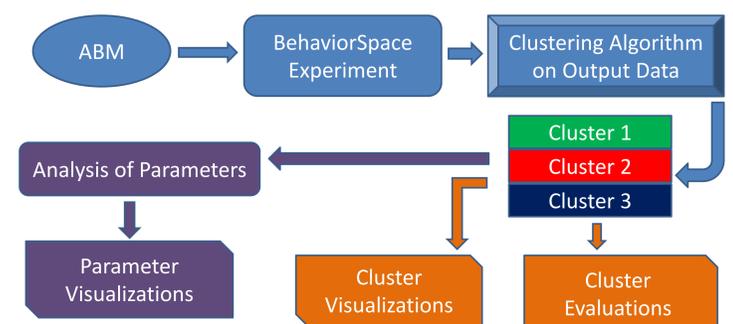
The flowchart above shows the process of evaluating agent-based models using unsupervised learning techniques. There are many opportunities for future work on this project. Ideally, this is an iterative process that can be refined to help a modeler build the most informative model and gain a full understanding of it. Research could be done into the best ways to adjust the parameters of the ABM based on the parameter distributions that appear in clusterings of the model's output data. Also, attempts could be made to automate some or all of this process to further reduce the amount of work for the researcher.

VI. References

[1] Wilensky, U., & Rand, W. (2015). An Introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo. Cambridge, Mass.: MIT Press.

[2] Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/ Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

[3] Sklearn.cluster.DBSCAN. Retrieved February 13, 2017, from http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

[4] 2.3. Clustering. Retrieved February 13, 2017, from http://scikit-learn.org/stable/modules/clustering.html