An Analysis of The Reverse Game of Life: Is The Chaotic Start of Life Predictable From Its Orderly Ends?

Junru Xiong
MSc Data Science
City, University of London

***Abstract*— The game of life is a particular type of cellular automaton through a set of rules, simulate the chaotic but patterned growth of biological organisms. Setting time backwards, predict the orderly starts with the end of chaos, which has received little attention. This research will investigate the analysis of random forest by regression and classification to implement in the Reverse Game of Life.**

Keywords—Game of Life, Cellular Automata, Random Forest, Regression, Classification

I. INTRODUCTION

Game of Life was invented in 1970 by mathematician John Conway. It is a special kind of cellular automata. Under simple rules, different initial states will evolve into various end states patterns. The game of life is played in a grids, each grid represents a cell. Cells are divided into "living cells" (denoted as 1) and "dead cells" (denoted as 0). Each cell has 8 "neighbours" grid around it. Game rules: If a dead cell has exactly 3 living neighbours, it becomes a living cell (birth); a living cell has 2 or 3 living neighbours, it continues to live (survive); otherwise, the cell dies or stay dead(Conway, 1970). Cells outside the grids are always dead.

The Reverse Game of Life aims to predict the start states given by end states. This research will explore this idea, based on the large amount of game data provided by Kaggle (2020), introduce to the random forest algorithm to critically analyse, and use different features and both regression and classification modelling outcome to evaluate, and deduce preliminary conclusion.

II. RESEARCH QUESTIONS AND ANALYTICAL APPROACH

A. Data and Domain

The dataset comes from Kaggle (2020) with 50000 training games in each 25×25 board. The foundation of Game of Life is Cellular Automata. Kumaresan and Gopalan (2017) have conducted in-depth research on the method of constructing symmetric cryptography by reversible cellular automata in the application of extracellular automaton cryptography. Due to its diverse evolutionary models, unexpected results have been achieved in terms of social self-organization mechanisms. It has also found application in various areas, including physics, cryptology, urban study, economics, and biology.

The dataset provided variables are (Kaggle, 2020):

The initial board was evolved 5 steps. The starting board's state was recorded after the 5 "warmup steps". These are the values in the start variables. The starting board was then evolved delta steps. Delta was chosen to be uniformly random between 1 and 5. If the stopping board was empty, the game was discarded.

Those variables contain:

- *id - unique identifier of each game*
- *delta - the number of steps between the start and stop boards*
- *start_0 - row 1, column 1 of the game's starting board*
- *start_1 - row 1, column 2 of the game's starting board*
- *...*
- *stop_0 - row 1, column 1 of the game's stopping board*

B. Research Question and Research Gap

This interesting game raises a novel question: If the game traces back into reversed order, is the chaotic start of Life predictable from its orderly ends? From the dataset, there are 25×25 boards which mean there are possible boards which is computationally impossible to use brute force solution. Moore (1962) studied whether the cellular automata have the "Garden of Eden", and proved that if a configuration has more than one precursor under certain conditions, then there must be a configuration without a precursor, which is so-called Garden of Eden pattern.

C. Analytical Approach

So, this issue has to move to machine learning and statistical methods. It will critically analyse and evaluate the implementation of machine learning methods in this problem. The overall aim of this paper is an experiment to see if machine learning could foresee the game of life in reverse.

III. ANALYSIS AND METHODOLOGY

A. Analysis

Firstly, it chooses 2 typical games visualization in Fig.1. This figure shows that the stop board cells position may somehow result in start board cells layout.

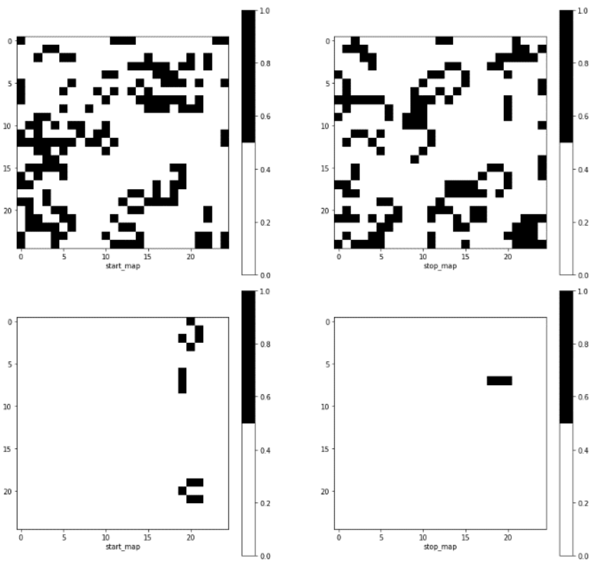

Figure.1 7th and 8th game star and end board visualization

Through the addition and subtraction of the start and end of the chessboard (Fig.2), it is a step further from the previous conjecture. Through observation, it can be found that the start and end cells positions of the board are not always random. They are either close or overlap.

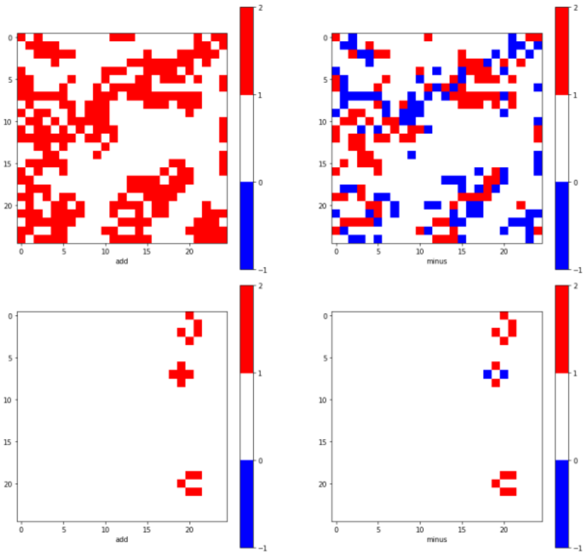

Figure.2 7th and 8th game star and end board cells addition and subtraction visualization

Moving to macro-scale observation, the following graph (Fig.3) shows the relationship between the sum of the stop and starts for each delta.

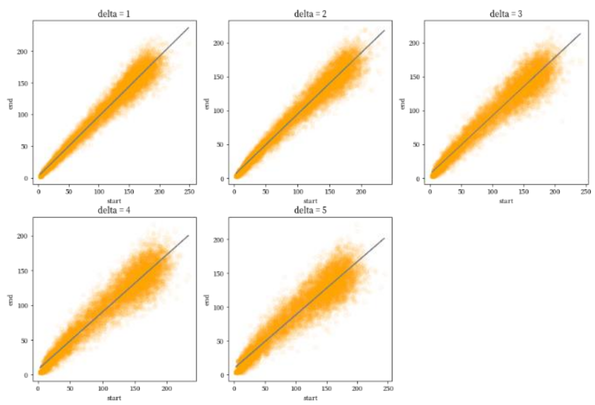

Figure.3 The relationship between the sum of the stop and starts for each delta

The relationship was observed to be relatively linear, and the more cells that survived, the more discrete their linear relationship became. As delta grows, the relationship between start and end becomes discrete.

Also, observing the regression line, it is found that the greater the delta, the less number of cells survive and that the range of variation is more significant as the delta increases. With the cells and delta increase, the start and end state living cells relationship getting discrete. So, It is hard to predict using white-box modelling.

B. *Methodology*

As the observations above, this research will introduce random forest classifier and regression to perform and evaluate their distinction.

Due to the Game of Life rule and data observation, each cell's death and survival is determined by its neighbours, and the end board cells position may either close or overlap. Therefore, it will use delta, the end board survival state of the cell (donate 1 as alive, 0 as death), and the number of neighbouring cells around each cell as the features. The predicted vector is the survival state of the cell in start board.

Firstly, it will directly use random forest classifier to predict the outcome, Although the accuracy score is 0.854, the predicted dataset is entirely 0. Random forest Classification is not well suited to learning the reverse game of life, due to how sparse the vectors are and training data is highly unbalanced (Fig.4). Passing in the original 25x25 live or dead values and the neighbours vector both fail to accurately learn the problem.

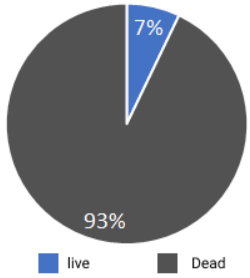

Figure.4 Total live and dead cells percentage in training dataset

Second, it attempts random forest regression, and converts the binary data of each cell into the survival probability. The output model's mean absolute error is 0.1215. Figure 5 selected two representative experimental results for visualization. It can observe that the converted probabilities relatively accurately predicted the approximate position from the end board to start board. However, the more number of cells and more dispersive on the training board, the more chaos the outcome will be.

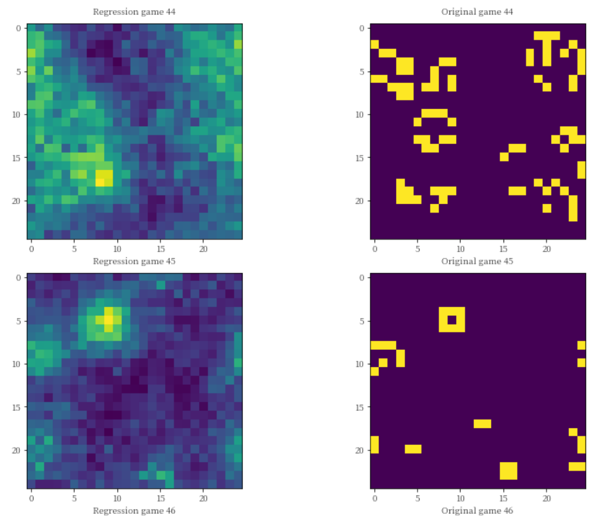

Figure.5 Comparison of regression random forest and original test dataset

Third, given results of the regression model have been obtained, is it possible to get accurate binary classification output by getting optimal threshold in regression output? Since the training data set is highly unbalanced, accuracy cannot be simply used to evaluate the results. AUC is significant metrics for evaluating classification models (Ling, Huang and Zhang, 2003). It will introduce the use of AUC rate to optimize the threshold.

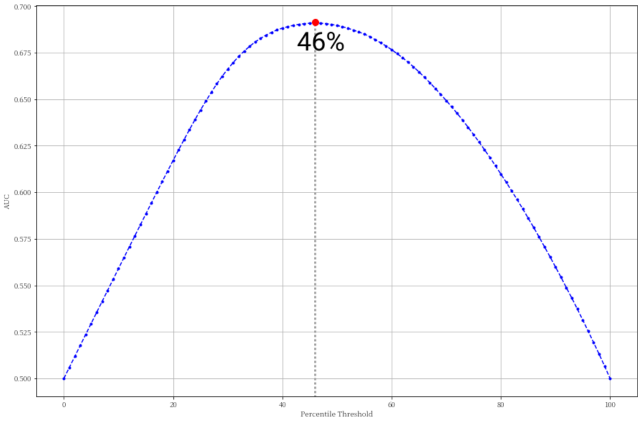

Figure.6 AUC threshold optimisation curve

Figure 6 shows each the iteration threshold from 1%-100%, and get the optimal threshold at 46%, and its AUC is 0.69.

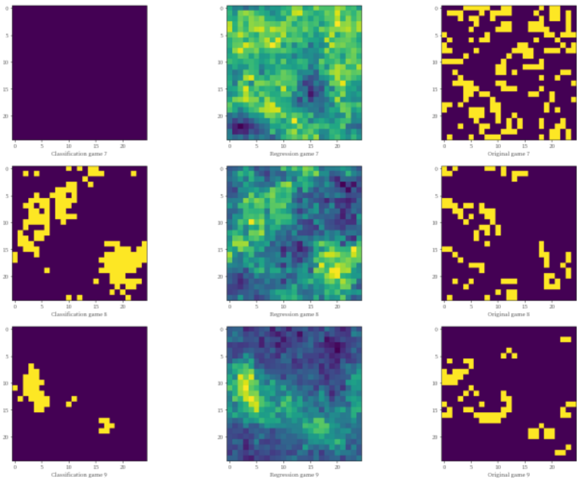

Figure.7 Comparison of optimized classification, regression random forest and original test dataset

Given the optimal classification model, Fig. 7 shows the comparison of optimized classification, random forest regression and original test dataset. It can be observed that although the optimized classification model might classify simple and relatively clustered input data correctly, it cannot handle noisy input data.

IV. Finding and Reflections

Through the above analysis, it then observes the overall distribution of prediction from a macro perspective.

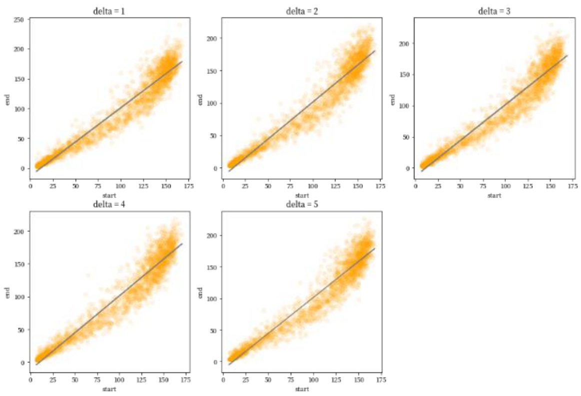

Figure.8 The relationship between the sum of the stop and starts for each delta by regression

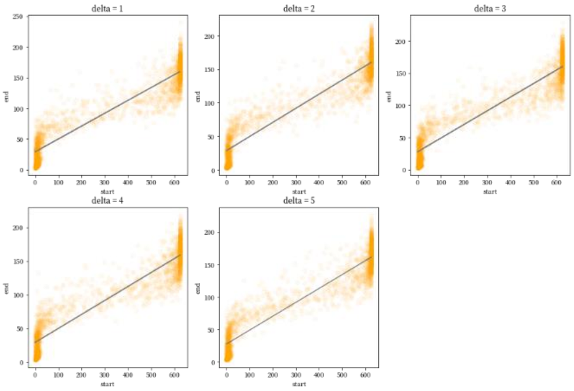

Figure.9 The relationship between the sum of the stop and starts for each delta by classification

Figures 8 and 9 indicate the relationship between the sum of stop and start in for each delta by regression and classification. From the scatter plot of the regression (Fig.8) that its relationship is relatively linear. Nevertheless, it does not become more discretized with the increase of delta. So, it may not able to predict the larger delta steps data. Similarly, the classification model does not discretize the linear relationship between start and end board as the delta increases. From Figure 9 that classification model output relationship is more extreme than regression, which does not match the distribution of the test dataset.

Therefore, after the above-analysed from micro individual board visualisation and macro scatter diagram relationship, the regression method performs well than classification when using the random to implement the reversed game of life. The poor performance of classification may get some loss due to the second time classification based on regression.

V. Conclusion and Futher Works

This research explored how to use the random forest to analyse and predict the reverse game of life. Machine learning methods may be able to predict the start board fuzzily, but as the delta grows, the relationship between its start and end board layout will become more discrete and more challenging to predict accurately. In the future research, the combination of interpretable mathematical methods and machine learning can be used, which may make predictions more accurate.

References

[1] Conway, J., 1970. The game of life. *Scientific American*, 223(4), p.4.

[2] Kaggle.com. 2020. Conway's Reverse Game Of Life 2020 | Kaggle. [online] Available at: <https://www.kaggle.com/c/conways-reverse-game-of-life-2020/discussion/200980> [Accessed 20 December 2020]..

[3] Kumaresan, G. and Gopalan, N.P., 2017. An Analytical Study of Cellular Automata and its Applications in Cryptography. *International Journal of Computer Network & Information Security*, 9(12).

[4] Ling, C.X., Huang, J. and Zhang, H., 2003, June. AUC: a better measure than accuracy in comparing learning algorithms. In *Conference of the canadian society for computational studies of intelligence* (pp. 329-341). Springer, Berlin, Heidelberg.

[5] Moore, E.F., 1962, December. Machine models of self-reproduction. *In Proceedings of symposia in applied mathematics* (Vol. 14, No. 5, pp. 17-33). American Mathematical Society New York