ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059

Approximating Chinese Character Pattern with Genetic Algorithm

Yaluo Wang

Foundations and Practice of Bio-Inspired Robotics New York, USA.

wangy296@newschool.edu

Abstract

The beauty of algorithms is a type of aesthetics that is beyond the emotional expression of the colors on paint brushes. In this paper, the research is accomplished through a coding program of Python to focus on approximating the selected Chinese character pattern, represented by two dimensional binary matrices, through generations of evolutions. The implementation of Chinese character is creative by combining technology with Chinese heritage, the character. How the technology is able to generate man-made characters through coding? The known steps of the Genetic Algorithm are the Initiation of population, Mutation, Selection and Crossover. Evolution is conducted on generations of populations, in which are candidate patterns, and the learning process can be tuned with multiple parameters including the mutation rate, the mutation fraction, the maximum generations and the population size. In addition, random immigrants are further introduced when the optimization score is not improving. When the process is successful, the result shows the best fitness score of 1.0, which implies 100% of similarity between the target Chinese character pattern and generated pattern. The algorithm is able to converge despite the grid size, and other setup of parameters including maximum generations, mutation rate, population numbers, etc. Above all, the rules of Game of Life are introduced as another intriguing approach for generating patterns on top of the Genetic Algorithm but should only be considered as an alternative.

Keywords

Genetic Algorithm; Game of Life; Chinese Character; Python.

1. Introduction

Recently, there is an upcoming trend to establish interdisciplinary practices between the art and design with computer science. As the advancement of technology, the image generation could come from computation and algorithms. By adopting the mathematical theories, one can manipulate the abstract parameters and create arts [1].

Initially, the goal was too ambitious to achieve, the primary target pattern was set up to a 300x300 size grid with over 1000 individuals in the population. Because of such a great population and grid size, the running time often takes hours to complete. The randomness in the pattern has also increased as the grid size becomes larger. The reason behind this is because there are more spaces for variations of cells to create patterns. Later in the process, the goal was altered into the following. The focus is on approximating the ideal Chinese character patterns after updating delta times with or without the Game of Life's rule. In order to find the fittest generated pattern compared to the target pattern, utilize the parameters such as populations and generations within the algorithm.

The issue, related to the randomness nature of the Genetic Algorithm when initializing the patterns population, can be solved with a large number of generations tested and having a higher mutation rate. The second issue that it may be difficult to achieve the best optimization towards the target

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059

pattern, resolve by monitoring the generation with limitations of smaller grids size and having the delta in the fitness function to be zero. The last problem with troubleshooting of the programs can be achieved with reference to the internet and other coding programs for assistance. Eventually, the result of 100% fitness score has achieved giving rise to the best optimization of the target pattern from the generated pattern. Automatically alter the parameters when the fitness score is fixed and end the process when the score stays the same beyond 800 generations. The second approach after adding the Game of Life theory; however, does not produce the same optimization due to much more constraints of the rules in the Game of Life.

In this paper, there are some basic lessons center around the successful implementation of genetic algorithms. The manipulation of the parameters should lead to a better result than before when the muted individual cells or the randomness of the population. The optimization of pattern is mostly depending on the delta in the fitness function compared to the other parameters in the process. Due to the constraint from Game of Life, the convergence is uncertain making the optimization harder or impossible to achieve.

With the above being discussed, the rest of the paper is organized as follows: In the section II, it is a reference to similar research and related work from other scholars. In Section III, the method of implementation is discussed in detail. In Section IV, it shows a specific presentation result in chosen methods. Lastly, Section V discusses the analysis of the final result and concludes the research as a whole. In the Section VI, there is appendix and after that is the reference in Section VII.

2. Related works

Recently, there is an upcoming trend to establish interdisciplinary practices between the art and design with computer science. Since the Game of Life and Genetic Algorithm are both theories that have been existed for a long time, there are already plenty of scholars' similar works in the field. To begin with, the Genetic Algorithm is based on the common Darwin's evolutionary theory that adopts into the computation [2]. The core message is to make slow changes for the best result through fitness function to evaluate individual solutions and populations' scoring through algorithm.

In addition, historically speaking, the Game of Life established by John H. Conway in the 1970s is a cellular automaton [3], which is a collection of fixed shape colored cells on a grid running with several specific rules. In the grid, each cell should have eight neighbors. If an occupied cell has either 0 or 1 neighbor, it is considered to be "loneliness", so the occupied cell dies by diminishing from the grid. If the occupied cell has over 4 to 8 neighbors around, it is considered to be "overcrowding", so it will also die. If the occupied cell has 2 or 3 neighbors, it survives meaning to stay in the grid for the next generation. If an unoccupied cell has 3 occupied neighbors, it will be born to the grid, occupied a new cell.

Scholars in the past have also tried to utilize the genetic algorithm to generate art patterns. Publication by the authors of University of Central Florida uses Genetic Algorithm to generate interesting patterns repeatedly with behaviors in Game of Life [4]. Moreover, there is the research of generating art tiles patterns with Genetic Algorithm [5]. The specific Genetic Algorithm implementation is reference to the work from the platform Towards Data Science, published author Pavel Tyshevskyi [6].

However, the revision was done in this specific research through multiple new elements. With the aim to have a better visualization, the animation was added in the coding program. The random immigrants are introduced to solve the potential fixed score after 400 generations, and the theory of Game of Life is built in the process of Selection of Genetic Algorithm as a second alternative [7-9].

3. Methods

The process starts by inputting the target pattern, which is a Chinese character meaning the king. The input pattern takes in the format of matrices, and then start the procedure of Genetic Algorithm. The specific steps include Selection, Mutation and Crossover. For reproduction the result, the parameters are shown in the Table 1.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059

Table 1. Example of Parameters Settings for Reproduction

Parameter	Value
Grid Size (start-up)	5x5 grid
Population Size (start-up)	100 individuals
Generation Size (start-up)	100 generations
Selection Methods	Fitness Function delta=0
Crossover Rate	Worth 1.0
Mutation Rate	0.05 (5% population)
Random Immigrants	0

The Genetic Algorithm has to have a range of population. Therefore, generate initial population of random individual solution with grids of the input pattern is necessary. After this, the Selection step keeps a fraction of population with the best fitness scores by calculating with Fitness Function and then randomly selected non-best individuals for the next generation. By using the Fitness Function to calculate the match rate, there are two approaches. The first approach is the basic step with delta=0 and the score generation happen when calculate match rates of the target character pattern with the given pattern.

On the other hand, with delta > 0, the fitness score is calculated differently. The match rate of the target pattern upon the given pattern is measured after delta times update through the rules of the Game of Life. The next step involves Mutation, in which randomly mutate a fraction of the current population by changing the cell values of selected patterns. During the process of Mutation, the rate is around 0.05 meaning only 5% of the population will generate mutation. Next, moving on to the step of Crossover, the birth of the two children pattern will take each gene from their parent patterns.

For innovation, the extra step of Random Immigrants that has random patterns introduced into the population within a low ratio. The important decision lies within the determination of the likeness by comparing the best score of generated patterns. If the score is almost equal to one, then the experiment of optimize the character patterns is considered to be successful. However, if the score is less than one meaning there is still a big difference between the target pattern with the generated pattern. With additional situation that the score is fixed in more than 400 generation, then increase the immigrant's ratio and increase the mutation rate. In another scenario when the generation is more than 800 generation but still fixed in the same score, the experiment has to put into stop as there is no use of doing more generations. Then, the experiment is considered as a failure but analyzation of the result could be useful for the research. (See in Appendix, Figure 6: Flowchart of the Method for Optimization the Pattern).

Generally, the fitness function is a type of evaluation that determines the closeness of a given solution to the optimized solution[10][11]. The fitness function commonly demonstrates as the inverse of the differences from the given to its optimization, and the result can determine which is the closeness of fitness. In order to calculate the fitness score for each generated pattern, the details of mathematics are shown as the following. Target Character Pattern Matrix subtraction with Generated Pattern Matrix:

$$\begin{pmatrix} 1, & 1, & 1, & 1, & 1 \\ 0, & 0, & 1, & 0, & 0 \\ 1, & 1, & 1, & 1, & 1 \\ 0, & 0, & 1, & 0, & 0 \\ 1, & 1, & 1, & 1, & 1 \end{pmatrix} - \begin{pmatrix} 1, & 1, & 1, & 1, & 0 \\ 1, & 0, & 1, & 0, & 1 \\ 1, & 1, & 0, & 1, & 1 \\ 1, & 0, & 1, & 0, & 1 \\ 1, & 1, & 1, & 1, & 1 \end{pmatrix} = \begin{pmatrix} 0, & 0, & 0, & 0, & 1 \\ -1, & 0, & 0, & 0, & -1 \\ 0, & 0, & 1, & 0, & 0, \\ -1, & 0, & 0, & 0, & -1 \\ 0, & 0, & 0, & 0, & 0, \end{pmatrix}$$

The result of the subtraction is the difference between the two patterns which should be positive, therefore:

$$|1| + |-1| + |-1| + |1| + |-1| + |-1| = 6$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059

Fitness calculates as $1 - \left[\frac{differencebetweenthetwopatterns}{gridsize}\right]$

thus, result in the score of the fitness:

$$1 - \left[\frac{6}{5 \times 5}\right] = 0.76$$

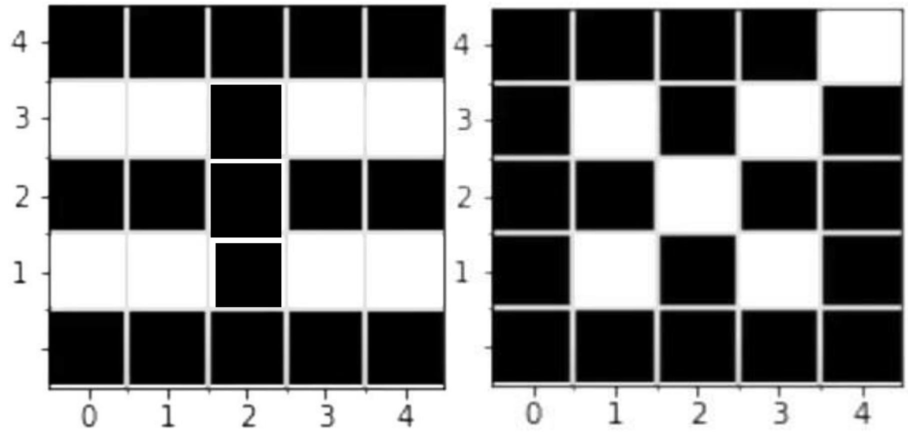


Figure 1. Target Character Patten with its Generated Pattern

4. Results

After the implementation of Genetic Algorithm to achieve the best solution, the result indicates several perspectives for analyzation. In the result, the prediction of 100% fitness and the best optimization was able to achieve (Shown in Figure 2).

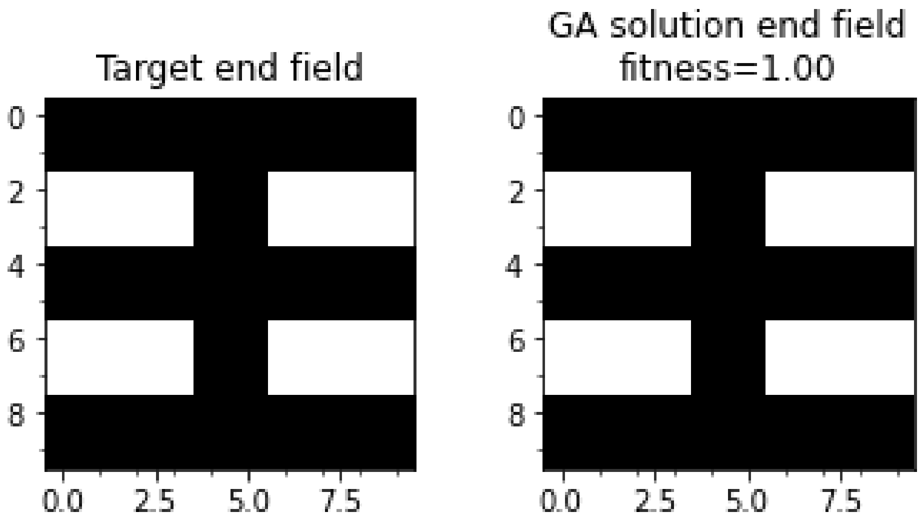
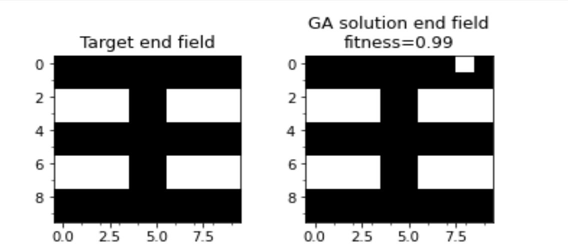
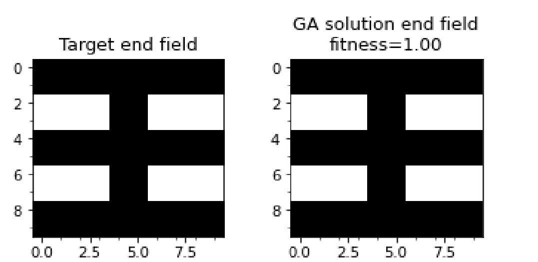


Figure 2. Comparison between Target Character Pattern with its Optimized Pattern

In running generations to produce new patterns with Genetic Algorithm, the problem of having fixated in one certain fitness score often appears, leading to no increase in the optimization. In order to solve such situation, the idea of adding the random immigrants should disrupt the population either increase of decrease the fitness score. The first experiment is restricting the delta, and making delta=0. The random immigrants help with generating the best result which shows details in the Table 2 below. After adding the random immigrants, the optimization could better achieve to 100 percentage. In contrast, without random immigrants, the population has variation and stop to generate the best score when reaches to 99 percentage. The introduction of immigrants is considered as efficient. [12]

Table 2. Generation Size Relationship with Top Fitness Score

Generation Size	Top Fitness Score	Top Fitness Score with Random Immigrants added after 300
#0-10	0.66	0.66
#10-20	0.67	0.67
#20-30	0.84	0.84
#30-40	0.90	0.90
#40-50	0.95	0.95
#50-60	0.97	0.97
#60-70	0.99	0.99
#70-300	0.99	0.99
#300-400	0.99 	1.0 

In the fitness function method within the Selection step of Genetic Algorithm, there is a choice in whether the delta is bigger than zero or equal to zero. When delta=0, without Game of Life Rules, the calculation of the fitness value could reach 100%, the match of the target pattern with individual patterns in the current population achieved optimization. In contemplation of advancement, with the Game of Life theory, it gives new perspective. When delta>0, calculate the match rates of the target pattern with the given pattern after delta Game of Life updates. Updating the pattern cells in the population after Game of Life updates, the optimization/best fitness score is harder or even impossible to achieve. The optimization with Genetic Algorithm in this research is better without the interruption of Game of Life rules.

As Figure 3 and 4 shows the relationship between the two variables: time and generations. Starting from the size of 10x10 Grid, the generation is set to be number in range from 100 to 1000, which is aiming to see how much time it would take by increasing the generation numbers. There is apparently a positive relationship between the Generation Size and Time to generate the best score. However, the outliers in the graph are still in debate with the questions of why they are existing. Figure 3 is with random immigrants and fitness measurement letting the programming to stop after reaching the best fitness score. In contrast, Figure 4 displays a relationship between time and generation without any additional restrictions. There is no introduction of random immigrants nor there is any pause in the generation. Compare Figure 3 and 4, it implies that the relationship between the generation and time is smoother and more linear without restrictions and addition individuals in the populations.

As shown in the Figure 5, the two variables of time and gird size have a relationship. The first graph is with certain innovations like introducing random immigration and constraining the fitness value; however, the second graph on the right does not have any innovation result in some outliers. In the generated experiment by comparing the data between the grid sizes with the generated time. In the case with innovation, the curve is smoother, which is different than the result from the previous relationship between the time used for best and updated generation. When increasing the grid size up to 300x300 with generation 1000 and population 500. When there is fewer innovation, the pattern generating becomes more and more turbulent as the grid size becomes larger. The outlier represents some of the uncertain cases after the grid size increase after 100 cells. The innovation helps to eliminate the outliers and uncertainty. Both of the graph represents an uprising exponential curve,

Volume 7 Issue 9, 2021
DOI: 10.6919/ICJE.202109_7(9).0059

which demonstrates a positive relationship between the time and grid size, slowly at first, but more rapid at last.

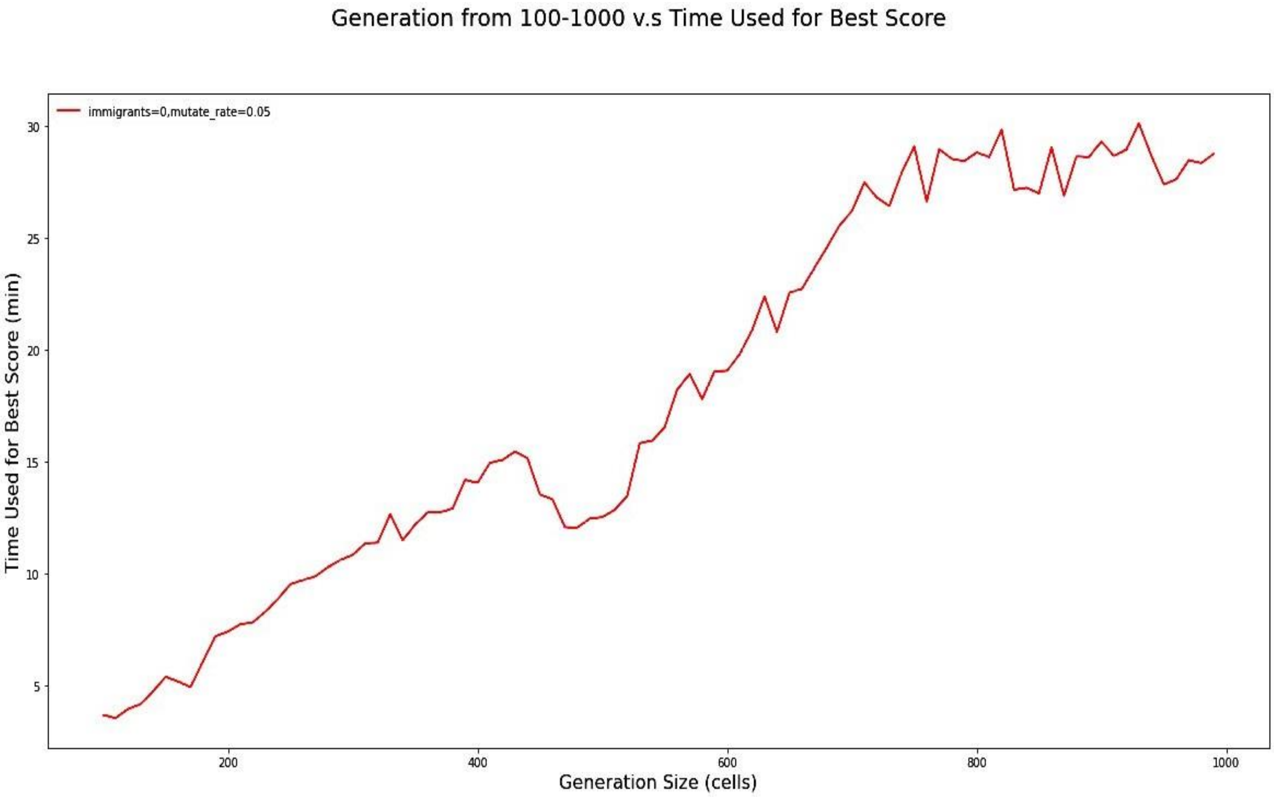


Figure 3. The relationship between the Time Used for Best Score v.s. Updated Generation from 100-1000 (with limitations in the fitness value and random immigrants)

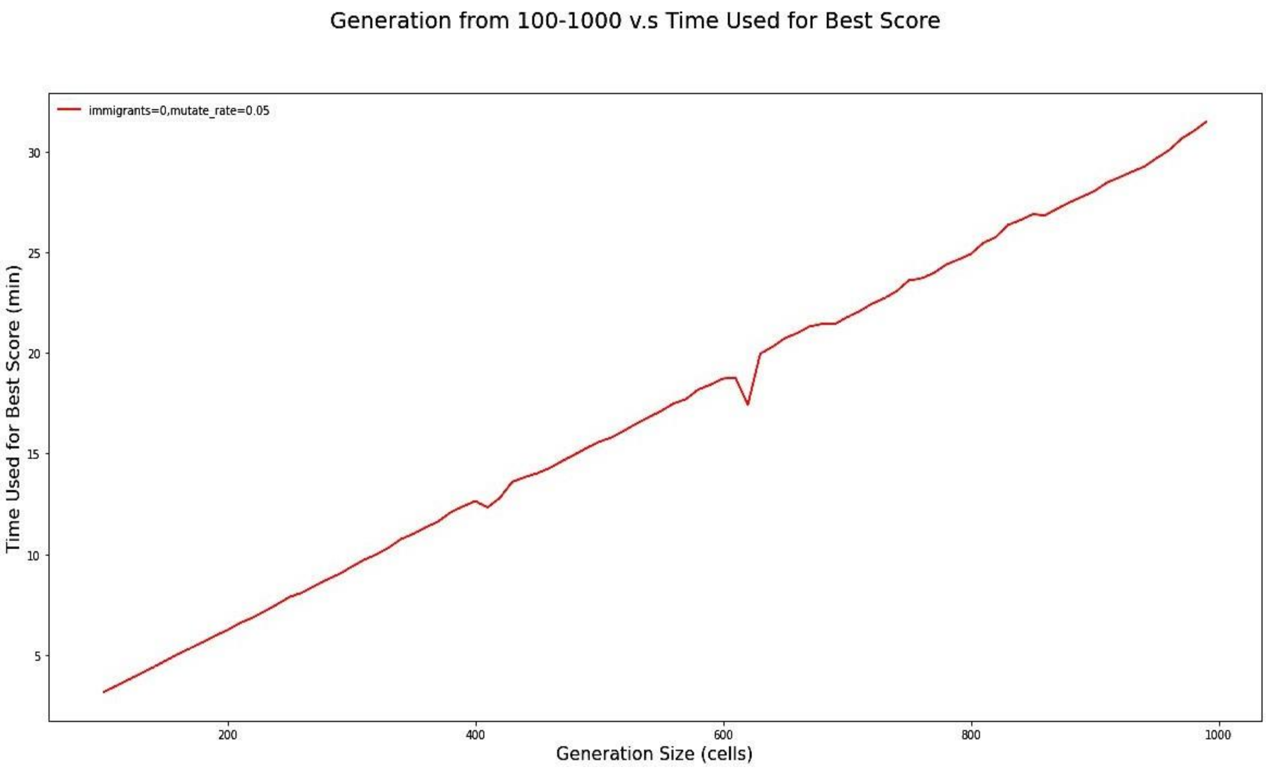
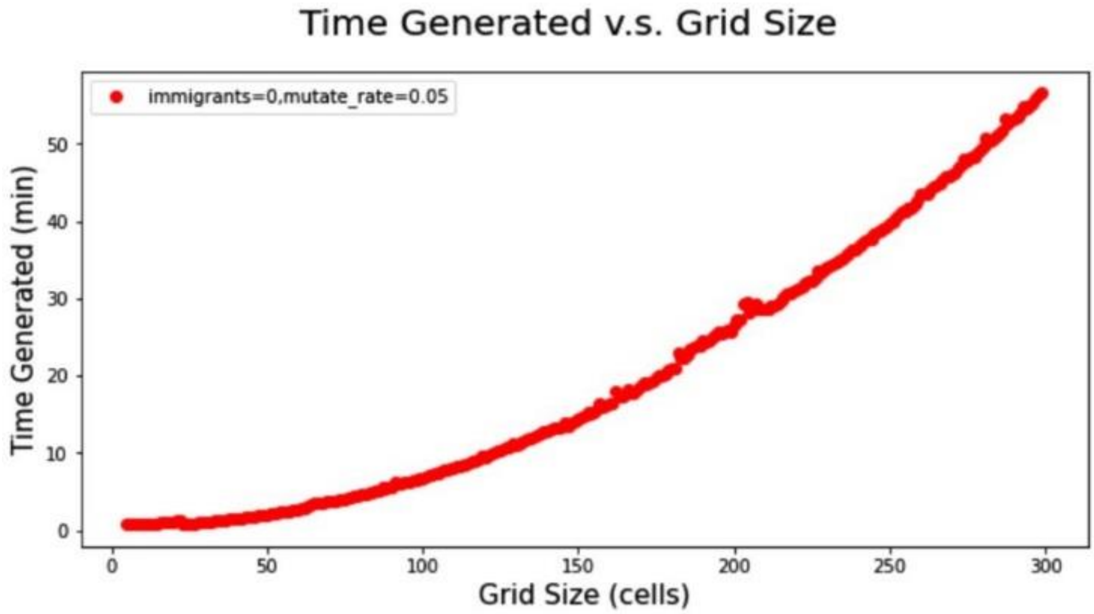


Figure 4. The relationship between the Time Used for Best Score v.s. Updated Generation from 100-1000 (without limitations nor random immigrants)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059



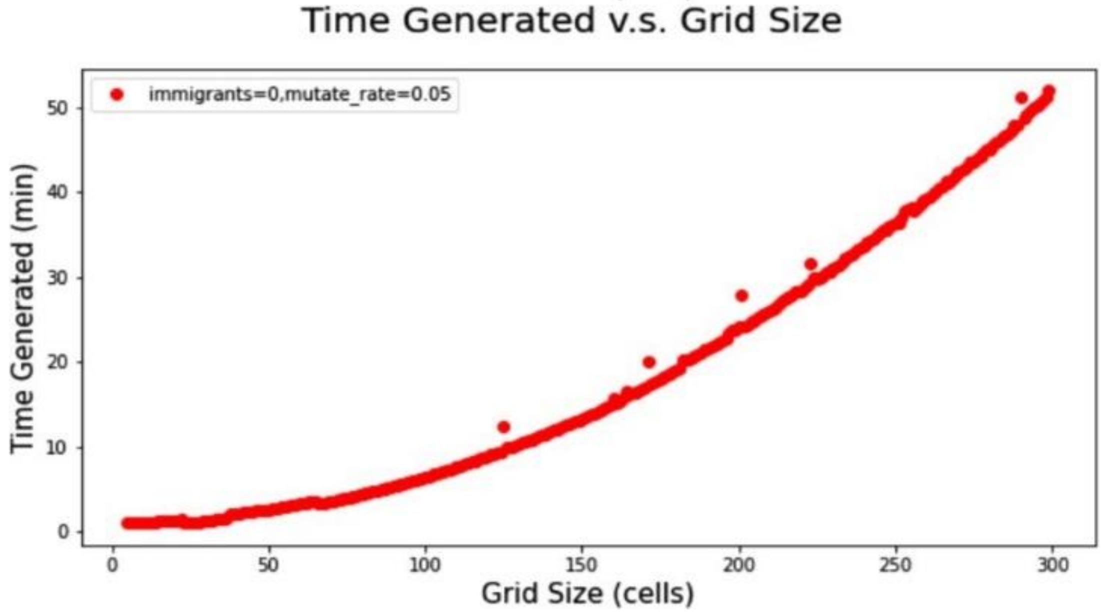


Figure 5. The relationship between the time Generated v.s Grid Size (left to right: with or without innovation)

5. Conclusion

In this paper, the implementation of the Genetic Algorithm is learned along with achieving the optimization of target pattern. To a certain extent, the research project is successful as best fitness score and optimization of 100% could be achieved. The predicted problems were able to be solved throughout the process. The analysis provides new insights of the possibility of Genetic Algorithm and the result after innovative approaches. The overall experiment is sufficient and illustrates relationship between different variables such as generation time, fitness scores and grid sizes. The introduction of random immigrants and constraint in generating patterns after best optimization, these innovations help to reduce outliers in the time generated when increasing the grid size. However, the innovations increase the uncertainty when only increasing the generation size. What could have done better is to eliminate the outliers with even larger generations and enhance the generated time to minimum. The next step of the research is to test out more Chinese character patterns for optimization and compare the different characters. Overall, the research project introduces a new perspective in the beauty of algorithms, which could inspire more interdisciplinary practice between art and computer science.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202109_7(9).0059

References

[1] Wolfram, S. (2002) A new kind of science. Wolfram Media.

[2] Eiben, A. E., & Smith, J. E. (2007) Introduction to evolutionary computing. Springer.

[3] Toffoli, T., & Margolus, N. (1993) A Cellular automata machines: a new environment for modeling. MIT Press,

[4] Alfaro, H. (1970) Generating Interesting Patterns in Conway ' s Game of Life Through a Genetic Algorithm Semantic Scholar.

[5] Heidarpour, M., & Hoseini, S. M. (2015) Generating art tile patterns using genetic algorithm. 2015 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS).

[6] P. Tyshevskyi. (2019) “Reversing Game of Life Using Genetic Algorithms,” Medium,. https://medium. com/ @ ptyshevs/rgol-ga-1cafc67db6c7 .

[7] J. VanderPlas, (2013) “Conway’s Game of Life in Python,” Conway's Game of Life in Python | Pythonic Perambulations. http://jakevdp.github.io/blog/2013/08/07/conways-game-of-life/.190898

[8] A. Gad, “Introduction to Optimization with Genetic Algorithm,” Medium, 03-Jul-2018. https:// towardsdatascience. com/introduction-to-optimization-with-genetic-algorithm-2f5001d9964b.

[9] J. Moukarzel, (2019) “From Scratch: The Game of Life,” Medium, https://towardsdatascience.com/from-scratch-the-game-of-life-161430453ee3.

[10] Arora, J.S., (2017) Introduction to Optimum Design (Fourth Edition).

[11] Mallawaarachchi, V. (2017) How To Define A Fitness Function In A Genetic Algorithm?. Medium. Available at: <https://towardsdatascience.com/how-to-define-a-fitness-function-in-a-genetic-algorithm-be572b9ea3b4>.

[12] Holden, S (2002) Python Web Programming. Sams Publishing; 1st edition.