View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by UWE Bristol Research Repository

Informatics in Medicine Unlocked 19 (2020) 100347

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Novelty search employed into the development of cancer treatment simulations

Michail-Antisthenis Tsompanas [a,*], Larry Bull [b], Andrew Adamatzky [a], Igor Balaz [c]

[a] Unconventional Computing Laboratory, University of the West of England, Bristol, BS16 1QY, UK
[b] Department of Computer Science and Creative Technologies, University of the West of England, Bristol, BS16 1QY, UK
[c] Laboratory for Meteorology, Physics and Biophysics, Faculty of Agriculture, Trg Dositeja Obradovica 8, University of Novi Sad, 21000, Novi Sad, Serbia

ARTICLE INFO

Keywords:
Novelty search
Cancer treatment
Evolutionary algorithm
PhysiCell
Simulator
Optimization

ABSTRACT

Conventional optimization methodologies may be hindered when the automated search is stuck into local optima because of a deceptive objective function landscape. Consequently, open ended search methodologies, such as novelty search, have been proposed to tackle this issue. Overlooking the objective, while putting pressure into discovering novel solutions may lead to better solutions in practical problems. Novelty search was employed here to optimize the simulated design of a targeted drug delivery system for tumor treatment under the PhysiCell simulator. A hybrid objective equation was used containing both the actual objective of an effective tumor treatment and the novelty measure of the possible solutions. Different weights of the two components of the hybrid equation were investigated to unveil the significance of each one.

1. Introduction

In optimization theory and machine learning, the idea of searching for possible solutions by putting more effort on the areas close to the optimum is well established. Nevertheless, these areas are determined based on an objective function that most definitely is riddled with local optima. It is logical that when the problem and, thus, its objective is complicated, the objective function will contain more local optima. A shortcoming of using solely an objective function is that areas in the search space, that may be stepping stones towards finding the global optimum, are neglected.

Novelty search [1] is an evolutionary search method that does not use as a guide the effectiveness of a solution, but its novelty measure. The novelty measure [9] is a parameter of the considered solution that indicates how far in the behaviour space is located compared with previously considered solutions. Novelty search was motivated by the fact that greedy search methods, which depend on a specific objective function, may suffer from deceptive evolution. Thus, the convergence to the optimum objective will in fact be hindered by this deception [1]. Moreover, novelty search was proposed to tackle the limited advance towards higher complexity that was observed when utilizing objective-based search methods based on objective functions.

Novelty search overlooks completely the objective, while it strives towards finding something new every time. Namely, the most novel behaviour that can be derived by utilizing each solution from the search space. The fact that multiple individuals merge to a single point in the behaviour space makes the methodology computationally viable. Moreover, as multiple solutions can merge to the same point in the behaviour space, it will be expected from novelty search to continue the search towards more complex solutions. Thus, it is expected to find a good solution in the way up the complexity ladder. This enables the mitigation of the concept of open-ended search (from simulated artificial life worlds) to real problems [2].

This methodology managed to outperform objective driven search in several real world problems. In the study where it was first suggested, an investigation on how to design artificial neural networks (ANNs), through neuro-evolution, that navigate a robot through a maze was performed [1]. Because of local optima in the objective space, namely dead-ends located close to the final target, the novelty search performed better than a greedy search. Risi et al. [3] implemented novelty search to a dynamic, reward-based single T-Maze problem. This kind of problem (and a couple of variations, like double T-Maze domain and a bee domain task studied in Ref. [3]) is equipped with an essential deceptive behaviour, that the novelty search managed to handle better than the well-established objective-based evolution.

PhysiCell [4] is a multicellular, agent-based simulator that was

* Corresponding author.
E-mail address: antisthenis.tsompanas@uwe.ac.uk (M.-A. Tsompanas).

https://doi.org/10.1016/j.imu.2020.100347
Received 23 March 2020; Received in revised form 4 May 2020; Accepted 4 May 2020
Available online 20 May 2020
2352-9148/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

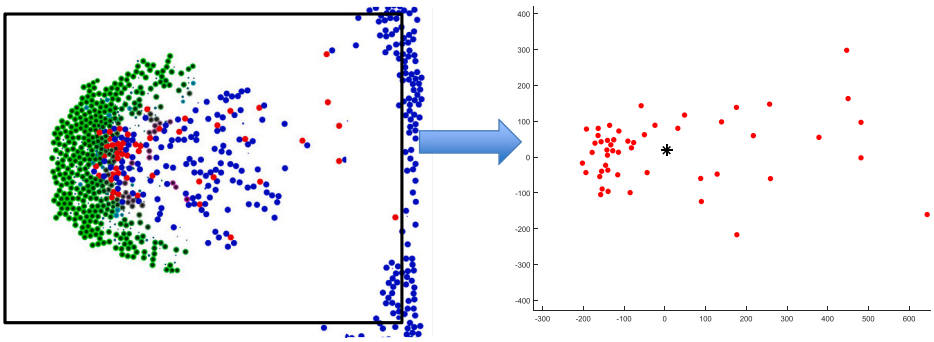

Fig. 1. Snapshot of simulation of PhysiCell [4] (left) and extraction of positioning of NPs (right). Red dots represent NPs, blue dots represent the therapeutic compound and green dots the cancer cells. The black star (right) is the average positioning of the ensemble of the NPs which is translated and used as the behaviour of the solution. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

designed to extend the BioFVM [5] framework, to form a virtual laboratory. PhysiCell is open source and offers several sample projects, one of which is studied here. More specifically, sample project "anti-cancer biorobots" [4] was developed as a possible tool to investigate the targeted cancer treatment, i.e. with drugs that adhere to specialized nanoparticles that would target specific molecules of the cancer cells.

The notion of PhysiCell serving as a guide to optimize the design of nanoparticle based cancer treatments [6,7] and discover cancer immunotherapies [8] was previously suggested. In the study of Preen et al. [6], PhysiCell was utilized to deliver surrogate-assisted evolutionary algorithms optimising the targeted delivery of a therapeutic compound to cancerous tumor cells. In the study of Tsompanas et al. [7] it was used under the same application of designing a therapeutic compound delivery system, as a target simulator for a new memetic algorithm, that is inspired by the fundamental haploid-diploid lifecycle of eukaryotic organisms. Finally, Ozik et al. [8] combined PhysiCell with active learning and genetic algorithms to dynamically probe a parameter space and unveil optimal cancer regression regions of immunotherapies.

2. Novelty search algorithm

The implementation of novelty search is possible by utilizing any evolutionary method, while changing the objective-based fitness function with a novelty measure [9]. As a result, this methodology compels the discovery of novel individuals. This new measure that will indicate how divergent each solution is compared with others in a behaviour space, should be defined based on the problem given. Choosing what the behaviour space will represent is not a priori obvious for every problem, as is the fitness function.

The novelty measure should represent how remotely located is the behaviour of every new individual, from the rest of the so far known ones, in the behaviour space. Thus, every new individual is compared with an archive of members of the previous generations in terms of their behaviour, and not their genotypes, to determine the new individual's novelty. This archive contains individuals that were considered novel during previous generations. A well-established measure to indicate that is sparseness [1], which can be defined as the average distance of the point under study with the k-nearest neighbors and given in the following equation.

$$\rho(x) = \frac{1}{k}\sum_{i=0}^{k} dist(x, \mu_i) \quad (1)$$

where ρ is the sparseness measure, k is the number of nearest neighbors considered, μ_i is the ith nearest neighbor, x is the individual under study and *dist* is a function returning the distance of the two points in the behaviour space. Consequently, the individuals that are located far away from clusters of others are assigned with higher sparseness values and, thus, are considered more novel.

Table 1
Unaltered parameters of PhysiCell simulator.

Parameter	Value
Damage rate	0.03333 min^{-1}
Repair rate	0.004167 min^{-1}
Drug death rate	0.004167 min^{-1}
Elastic coefficient	0.05 min^{-1}
Cargo O_2 relative uptake	0.1 min^{-1}
Cargo apoptosis rate	4.065e-5 min^{-1}
Cargo relative adhesion	0
Cargo relative repulsion	5
Maximum relative cell adhesion distance	1.25
Maximum elastic displacement	50 μm
Maximum attachment distance	18 μm
Minimum attachment distance	14 μm
Motility shutdown detection threshold	0.001
Attachment receptor threshold	0.1
Worker migration speed	2 $\mu m/min$
Worker apoptosis rate	0 min^{-1}
Worker O_2 relative uptake	0.1 min^{-1}

Similar to a conventional fitness function space of a real problem, behaviour space can not be perfectly mapped beforehand its investigation by the evolutionary methodology. As a result, the novel individuals can be discovered only through an exploration procedure, analogous to locating the areas close to optima of the conventional objective. Moreover, novelty search has an inherent coevolutionary nature, given that the sparseness is calculated as a distance from previously discovered novel individuals throughout the evolution process.

When a new individual has a comparably large sparseness/novelty measure value, meaning it is novel in the present generation, it is added in the aforementioned archive of novel individuals. Therefore, having this archive as a guide of where the search procedure have already sought for solutions, the methodology strives towards unexposed areas of the behaviour space, most probably containing more complex and better solutions.

3. Methodology

The optimization of the parameter set that determines the efficiency of nanoparticles in a cancer treatment, defined as worker agents in simulator PhysiCell (v.1.5.1) [4] was investigated here. Namely, sample project "anti-cancer biorobots" [4] was utilized to simulate the effect of different kinds of nanoparticles. For instance, the result of a simulation is demonstrated in Fig. 1, where blue dots depict the therapeutic compound that is deposited by the nanoparticles (red dots) close to the cancer cells (green dots). Cancer cells decay and eventually die

(different shades of green color dots in Fig. 1), due to the proximity of the therapeutic compound (for more details refer to Ref. [4]).

As population based optimization methodologies were used, individuals were defined in a 6-D space of possible combinations of the simulated nanoparticle parameters. These parameters along with their ranges are: attached worker migration bias [0,1], unattached worker migration bias [0,1], worker relative adhesion [0,10], worker motility persistence time (in *mins*) [0,10], worker relative repulsion [0,10] and the cargo release O_2 threshold (in *mmHg*) [0,20]. All other parameters of the simulator were not modified throughout the evolutionary process and set at values same as in the initial distribution of the simulator (PhysiCell v.1.5.1 [4]), illustrated in Table 1.

The scenario of the sample project "anti-cancer biorobots" involves the initialization of a 200μm radius collection of cancer cells (approximately 570 simulation agents) which then grows for 7 days by duplicating randomly chosen cancer cells. Then the therapy is applied for 3 additional days. The therapy is comprised by 50 simulation agents representing nanoparticles and 450 simulation agents representing the therapeutic compound.

To alleviate a part of the effect of the stochastic nature of the simulator on the results, a single tumor was used for testing every possible individual in the search space. The aforementioned tumor was produced after evolving in the simulator an initial 200μm radius collection of cancer cells for a simulated period of 7 days. Then, for each test the fully grown tumor was loaded to the simulator (after changes in the initial source code) and the treatment was applied immediately. The test was finalized after 3 days from the introduction of the treatment (an example is illustrated in the left part of Fig. 1), namely a total simulation time of 10 days from initial 200 μm radius tumor. Nonetheless, to further minimize the effect of the stochastic procedure, the average of the outputs after 5 runs of the simulator with the same set of parameters was examined. The objective fitness of each solution was determined as the remaining cancer cells in the simulated area after the 3 days of simulated treatment. Note that the execution time for testing each possible solution (5 runs of the simulator with the same parameters and averaging results), on an Intel® Xeon® CPU E5-2650 (using 8 of the 48 cores) at 2.20 GHz with 64 GB RAM requires approximately 6 min of wall-clock time.

As a reference point, the optimization of the worker agents of PhysiCell was attempted by a generic genetic algorithm (GA). The population of the GA was of size $P = 20$. The tournament method was used for parents' selection and replacement by mutated offspring with size $T = 2$. Moreover, uniform crossover with probability $X = 80\%$ was implemented and mutation rate per allele of $\mu = 20\%$ with random step size of $s = [-5, 5]\%$. Note that the population was evolved in generations (here for 10 generations), namely all individuals from the previous population were compared with the offspring and replaced appropriately to form the next generation.

For the proposed methodology of novelty search, the same algorithm as in the aforementioned was used, whereas the fitness function was altered to incorporate the novelty measure. It is noteworthy that despite the fact that the population size is not substantially large, it is chosen due to a limited computational load budget. Moreover, using the same population size for both algorithms alleviates any effects that the trade-off between accuracy and execution time may impose. It is suggested that novelty search can be implemented in hybrid fitness functions, using both novelty measure and the objective [10,11]. Using that as a motivation, we designated a hybrid fitness function as in the following:

$$fitness = \frac{rcc}{rcc_{thr}} - \frac{sparseness}{s_{thr}} \quad (2)$$

where *rcc* is the remaining number of cancer cells after the 3 days of the cancer treatment and *sparseness* the average distance of the new individual's behaviour from the 5 nearest neighbors in the behaviour space (as defined in Eq. (1)). Moreover, rcc_{thr} is a weighted parameter to normalize the values of the remaining number of cancer cells. The

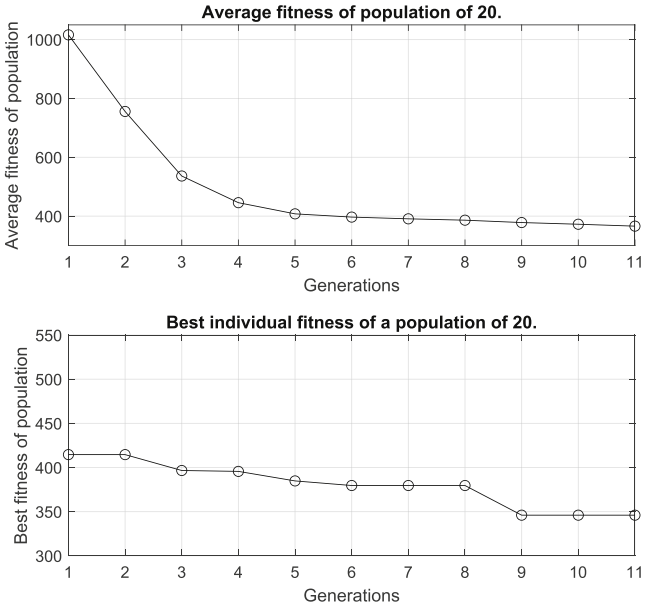

Fig. 2. Average and best actual fitness of individuals in each generation for the simple GA.

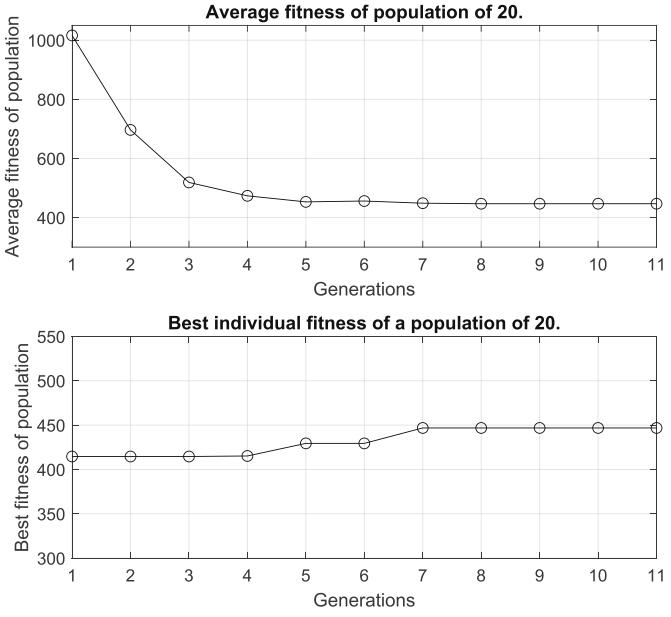

Fig. 3. Average and best actual fitness of individuals in each generation for the hybrid fitness function with $s_{thr} = 200$.

assigned value is equal to 1400 for the following experiments, in order to normalize the first part of the fitness function in a range of $[0, 1]$. Finally, s_{thr} is a weighted parameter to normalize the values of the sparseness (or novelty measure). For the following experiments this parameter is taking values in the range of 200–1000 (with intervals of 200), in order to normalize similarly the second part of the fitness function in a range of $[0, 1]$. The higher values of s_{thr} obviously result to weaker effects of the novelty measure to the hybrid fitness function.

The output of each solution in the behaviour space was defined as the center of gravity of the ensemble of worker agents at their final position after 3 days of simulated treatment. More specifically, the placement of the collection of nanoparticles in the simulated area (an example is depicted in the right part of Fig. 1 as a black star). This behaviour is easily calculated by the average of the coordinates of all the worker

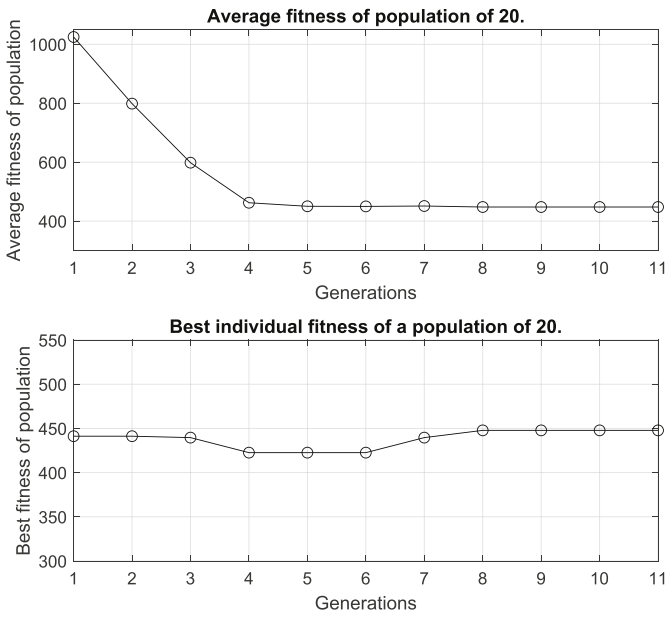

Fig. 4. Average and best actual fitness of individuals in each generation for the hybrid fitness function with $s_{thr} = 400$.

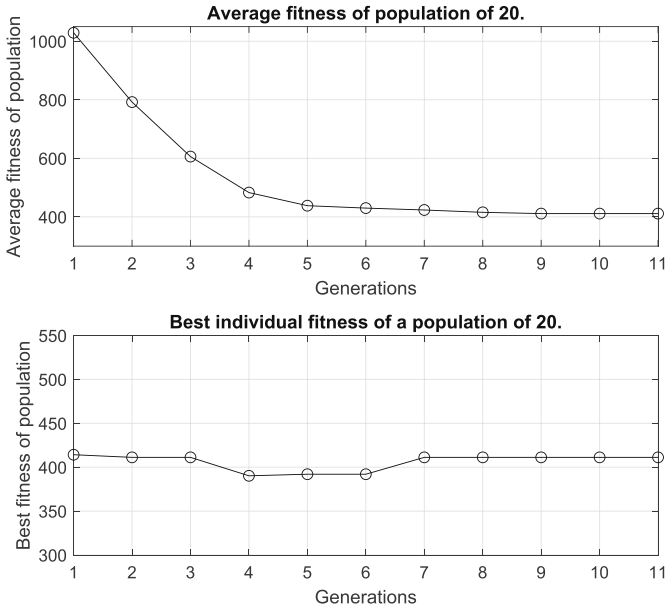

Fig. 5. Average and best actual fitness of individuals in each generation for the hybrid fitness function with $s_{thr} = 600$.

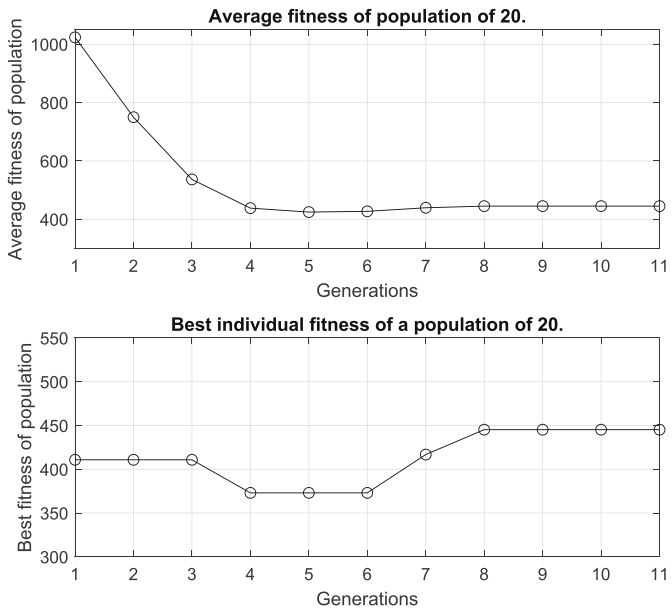

Fig. 6. Average and best actual fitness of individuals in each generation for the hybrid fitness function with $s_{thr} = 800$.

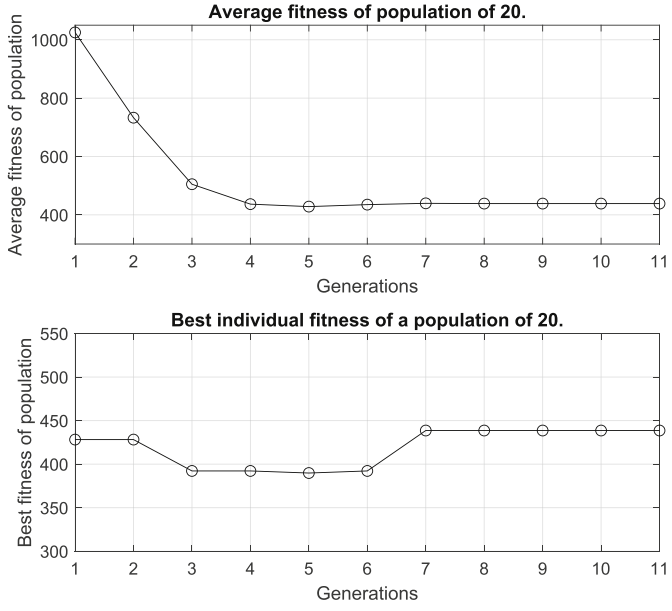

Fig. 7. Average and best actual fitness of individuals in each generation for the hybrid fitness function with $s_{thr} = 1000$.

agents. Consequently, as in previous works of novelty search [1,12], the topology of the result was taken into account, which is ignoring the actual objective.

4. Results

To make the comparison between different algorithms meaningful the initial population for every case is composed by the same individuals. Three different sets (of $P = 20$ individuals) of initial populations were tested. The outputs of using a generic GA and the hybrid fitness function (as described in Eq. (2)) with different normalization parameter s_{thr} are depicted in Figs. 2–7. These figures illustrate the average actual fitness of the population for every generation and the actual fitness of the best individual found in each generation. By the term actual fitness, we define the number of remaining cancer cells in the simulated area, not to be confused with the hybrid fitness function used in novelty search method and given in Eq. (2).

From the results in Figs. 2–7 it is established that while the simple GA provides a better (or at least the same) fitness for every generation, the novelty search method presents a more erratic behaviour. Namely, with $s_{thr} = 200$ it does not manage to find a better solution than the initial randomly generated one, on the contrary it searches the solution landscape without any profound advance in fitness. However, for higher values of the parameter s_{thr} (meaning smaller significance of the novelty measure compared with the actual fitness), the searching method manages to optimize the solution at least briefly in the extend of the 10 generations. More specifically, as depicted in Figs. 4–7 there is a decline in the amount of remaining cancer cells (actual fitness) for the up to the

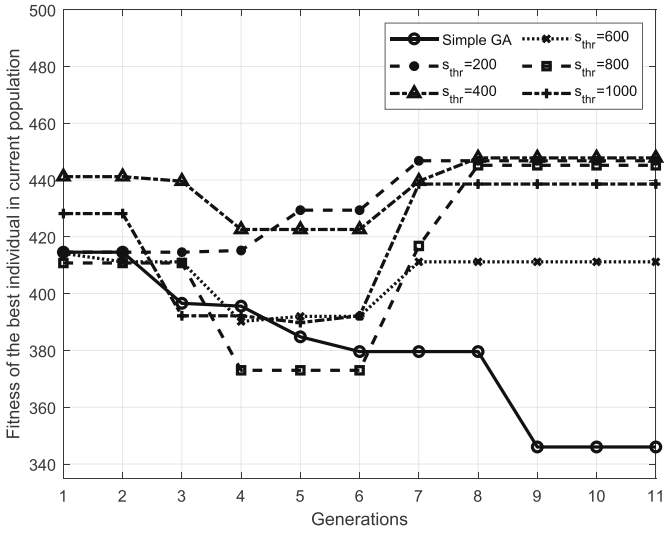

Fig. 8. Cumulative results of first run. Best actual fitness of individuals in each generation for all s_{thr} parameters compared with the simple GA.

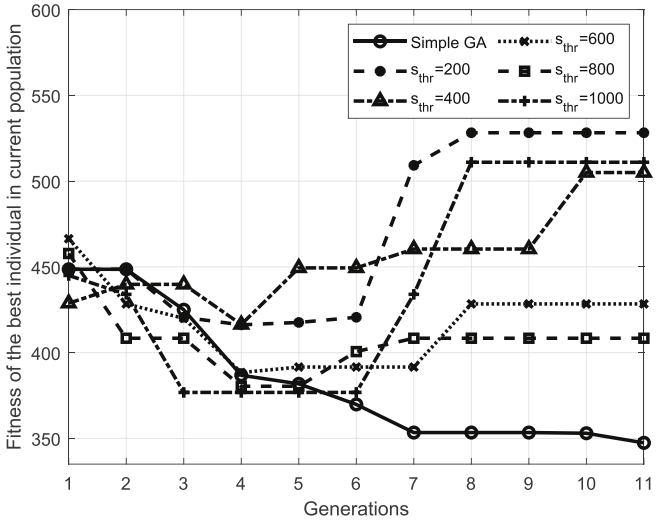

Fig. 9. Cumulative results of second run. Best actual fitness of individuals in each generation for all s_{thr} parameters compared with the simple GA.

6th generation, but then the novelty measure seems to be putting more pressure into finding more novel solutions than remaining the fittest in the population. The decline in the actual fitness is more profound for the middle values in the range of the s_{thr} parameter, specifically for $s_{thr} = 800$.

To better compare the results of the search method with different s_{thr} parameters and the simple GA, Figs. 8–10 are provided. Each figure is containing the results of every run, namely the use of different methods in the same initial population of $P = 20$ individuals.

Throughout all the different runs, it can be observed that novelty search yields more erratic outputs when studying the actual fitness. An outcome that is expected, given the fact that the hybrid fitness function used in this search method contains the novelty measure that completely ignores the actual fitness of the solutions. Nevertheless, it can be noticed that while in the final results (after 10 generations of artificial evolution) simple GA is providing better solutions, in most of the initial generations, novelty search is providing better solutions. In particular for s_{thr} parameters higher than 400.

This fact is better illustrated in Fig. 11. Here the best individual

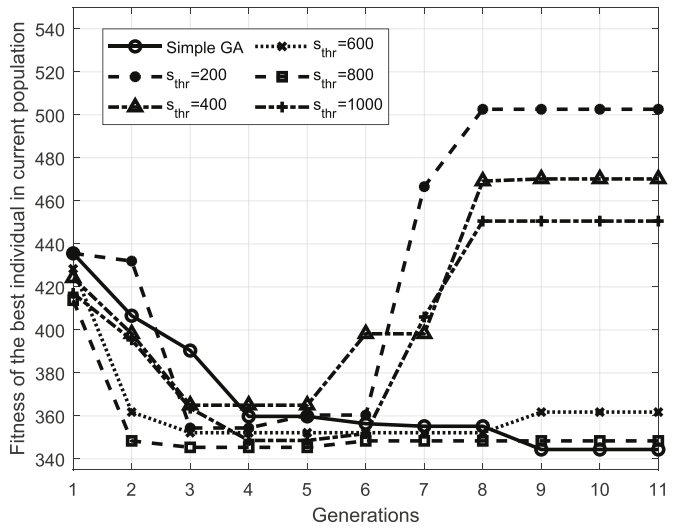

Fig. 10. Cumulative results of third run. Best actual fitness of individuals in each generation for all s_{thr} parameters compared with the simple GA.

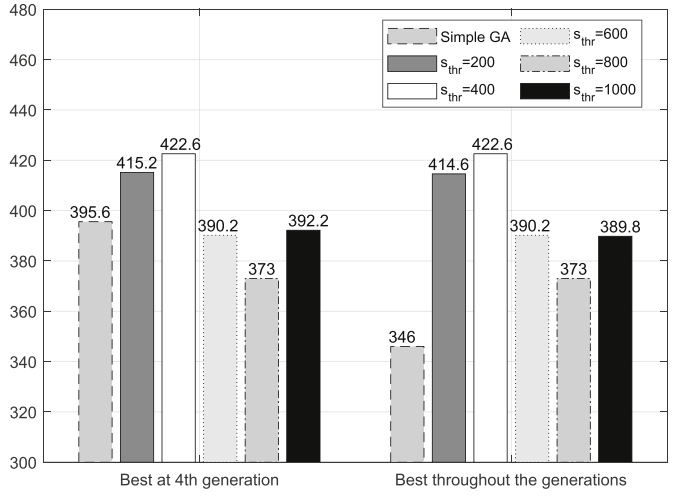

Fig. 11. Best actual fitness of individuals in 4th generation and throughout all generations.

discovered until the 4th generation and throughout all the generations is presented for the simple GA and the different cases of the hybrid novelty search for the first run. Despite the fact that the simple GA seems to outperform the novelty search throughout the 10 generations, it seems that the novelty search with $s_{thr} \geq 600$ outperforms the simple GA for up to the 4th generation.

The same finding stands for all three runs (different initialization of the comparison test). This can be realized by Figs. 12 and 13 rendering the boxplots of the best individual in terms of actual fitness up until the 4th generation and throughout the length of the all the generations.

5. Conclusion

Novelty search is motivated by the need to overcome the problems of deception and local optima inherent in objective optimization. Ignoring the objective completely or using hybrid fitness functions including a novelty measure, may often benefit the search of a better solution. In this study, this methodology was employed to optimize the design of targeted drug delivery systems, aiming cancerous tumours. The solutions were evaluated by PhysiCell simulator, namely by its sample project "anti-cancer biorobots". While PhysiCell [4] was previously studied

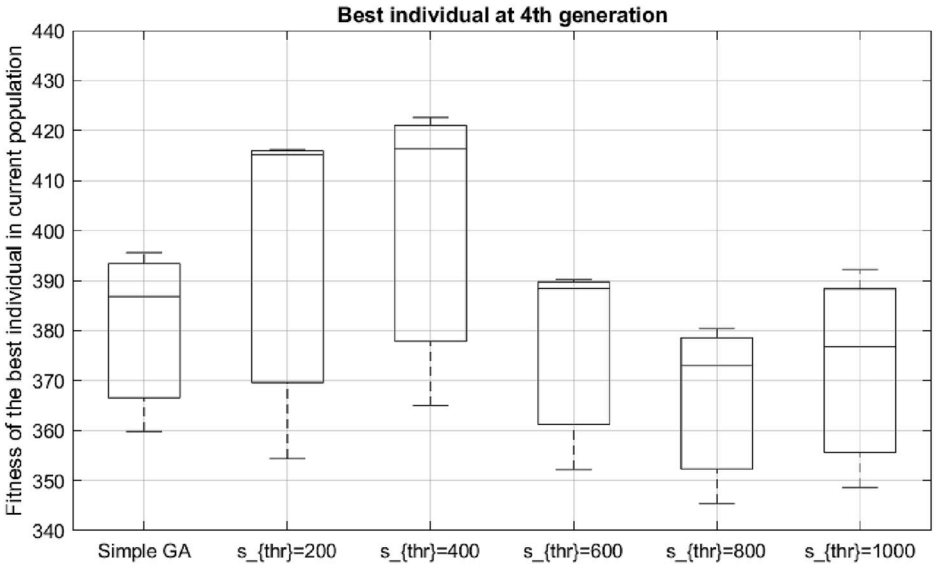

Fig. 12. Boxplot of the best actual fitness of individuals in 4th generation for all runs.

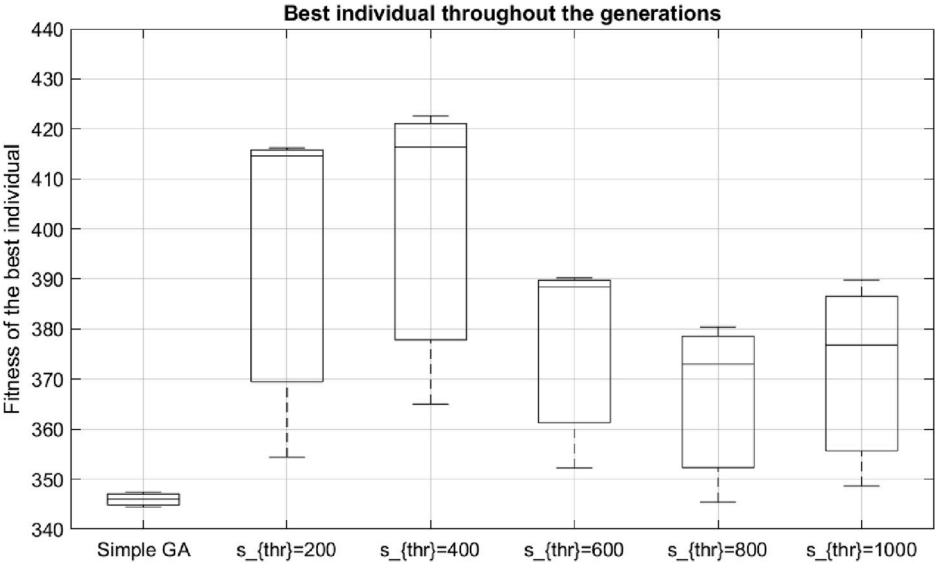

Fig. 13. Boxplot of the best actual fitness of individuals in all generations for all runs.

under different optimization algorithms [6–8] that were always guided by conventional objectives, here the evolutionary search of solutions is guided by including the novelty measure of these solutions considered in their behaviour space.

The association of the fitness function with a novelty measure rather than only the objective proved to lead to more efficient solutions faster in the initial steps of artificial evolution. Moreover, an analysis of the significance of the novelty measure was performed by running optimization processes with different weights on the novelty measure. The medium and high values in the range studied proved to be more effective.

Nonetheless, novelty search has some limitations. Given the fact that it ignores the objective, there is no pressure towards further optimization once a good solution is found, which is not ideal. An optimized solution may be produced by novelty search only if an individual can appear novel, while demonstrating this optimized performance. As illustrated in the results provided, a simple GA was able to outperform the hybrid novelty search in the course of 10 evolution steps. A possible solution to this limitation, is to take the most promising results from novelty search and further optimize them based on an objective function. Thus, following this procedure will take advantage of the strengths of both approaches. Novelty search successfully locates the approximate solutions, while objective optimization further investigates the local area around approximate solutions.

On the other hand, novelty search can be applied in the case where a traditional evolutionary algorithm reaches convergence, to inject the population with new, diversified individuals. These prospects of combined novelty and objective based procedures can serve as aspects of future work. Finally, the conclusions driven from this study, will be applied on ongoing research [6,7] towards a more wide applicability platform that will design, develop and evaluate drug delivery systems aiming cancer tumours. Namely, using hybrid fitness functions and novelty measures in the application of surrogate-assisted evolutionary algorithms [6] or unconventional, innovative evolutionary techniques, like haploid-diploid methodology [7], could help avoid being stuck in potential local optima.

Declaration of competing interest

None Declared.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 FET Open programme under grant agreement No. 800983.

References

[1] Lehman J, Stanley KO. Exploiting open-endedness to solve problems through the search for novelty. Artif Life 2008;11:329.
[2] Lehman J, Stanley KO. Revising the evolutionary computation abstraction: minimal criteria novelty search. In: Proceedings of the 12th annual conference on Genetic and evolutionary computation; 2010. p. 103–10.
[3] Risi S, Hughes CE, Stanley KO. Evolving plastic neural networks with novelty search. Adapt Behav 2010;18(6):470–91.
[4] Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. Physicell: an open source physics-based cell simulator for 3-d multicellular systems. PLoS Comput Biol 2018;14(2):e1005991.
[5] Ghaffarizadeh A, Friedman SH, Macklin P. Biofvm: an efficient, parallelized diffusive transport solver for 3-d biological simulations. Bioinformatics 2015;32(8):1256–8.
[6] Preen RJ, Bull L, Adamatzky A. Towards an evolvable cancer treatment simulator. Biosystems 2019;182:1–7. https://doi.org/10.1016/j.biosystems.2019.05.005. http://www.sciencedirect.com/science/article/pii/S0303264719300206.
[7] Tsompanas M-A, Bull L, Adamatzky A, Balaz I. Haploid-diploid evolution: nature's memetic algorithm. 2019, 07302. arXiv:1911.
[8] Ozik J, Collier N, Heiland R, An G, Macklin P. Learning-accelerated discovery of immune-tumour interactions. Molecular Systems Design & Engineering; 2019.
[9] Lehman J, Stanley KO. Efficiently evolving programs through the search for novelty. In: Proceedings of the 12th annual conference on Genetic and evolutionary computation; 2010. p. 837–44.
[10] Stanley KO, Lehman J. Why greatness cannot be planned: the myth of the objective. Springer; 2015.
[11] Mouret J-B. Novelty-based multiobjectivization. In: Doncieux S, Bredèche N, Mouret J-B, editors. New horizons in evolutionary robotics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 139–54.
[12] Gomes J, Urbano P, Christensen AL. Evolution of swarm robotics systems with novelty search. Swarm Intelligence 2013;7(2):115–44. https://doi.org/10.1007/s11721-013-0081-z.