International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 8 (2019), pp. 1183-1190

Bioinspired Optimization Tool for the Investment Portfolio Selection Problem

Holman Montiel Ariza[1], Harvey Gómez Castillo[2] and Luz A. Gaviria Roa[3]

[1,2] *Facultad Tecnológica, Universidad Distrital Francisco José de Caldas, Bogotá D.C, Colombia.*

[1]*ORCID:* 0000-0002-6077-3510, [2]*ORCID:* 0000-0003-4474-3642

[3] *Facultad de Ingeniería, Fundación Universitaria Panamericana, Bogotá D.C., Colombia.*

Abstract

An investment portfolio can have many configurations, since its composition varies depending on the investor's capital. This capital gives the portfolio growth potential, which depends on the companies, sectors or materials in which an initial investment is made. Bearing in mind that there are many investment sectors, finding an appropriate configuration that allows the portfolio to maximize the profits obtained by the investor remains a challenge. Therefore, this article proposes a bio-inspired optimization algorithm that allows the user to maximize the profits of an investment, through a computational model that estimates the performance of a group of investment sectors selected in a stochastic manner. One of the advantages of the proposed algorithm is that it maximizes the portfolio's profits by relying on real data from an investment exchange.

Keywords: Optimization Algorithm, Stochastic, Bio-inspired, Genetic Algorithm, Computational Model.

1. INTRODUCTION

An Investment Portfolio (IP) or securities portfolio is a set of instruments or Financial Assets (FA) that an investor may have, for example; shares in a company, foreign currency, art, securities, among others. Considering the risk involved in forming an IP, financial institutions propose different options to investors depending on their risk profile. The risk profile determines an investor's likelihood of achieving his or her objectives at the time the investment is made and is classified in two non-systematic and systematic ways [1-3].

Non-systematic risk groups all operations in which the probability of not achieving the proposed objectives is minimized or controlled, for example, when a manager increases the company's profits by increasing sales. Unlike nonsystematic risk, systematic risk is a factor that is always present in investment portfolios and cannot be eliminated, for example, interest rates, taxes and economic recession. In both cases, there is the possibility of not achieving the goals, therefore, some authors have developed techniques based on the behavior of FA to optimize the utilities of an IP [1-3].

One of the techniques commonly used by financial institutions is diversification, which consists of offering the investor a set of assets so that he has several options to increase the return on his profits. The disadvantage of this technique is that profits grow relatively slowly compared to other IP optimization techniques. However, it offers confidence to the investor, since it reduces the risk of not achieving the established goals. Other techniques are based on the construction of histories that give the investor a view of the behavior of stocks during certain time intervals. Normally, this technique is used by brokers to determine the feasibility of investing in certain FA. Unlike diversified techniques, the techniques or methodologies adopted by brokers increase the risk of not achieving IP goals, but profits are growing relatively fast [4, 5].

This topic has also been addressed in the research field by trying to optimize the profits of the investment portfolio through techniques that implement mathematical models or software-based applications. On the one hand, mathematical models tend to focus on certain areas of PI, since there is a large amount of FA in which money can be invested. For example, there are models for manufacturing industry, oil, gas or metals. On the other hand, software-based applications present a realtime trend graph of the value of FA and give investors the option to invest or withdraw their assets quickly [6-10].

As can be seen, this topic is still under development due to the amount of information that must be known and interpreted, which through current models or techniques does not allow the investor to properly project an operation in the future [9]. Because, the projection is done by a manual process, which adds some margin of error. Therefore, this paper proposes a technique to automate the FA selection process for the formation of a PI. This technique is based on a bio-inspired optimization algorithm that maximizes the performance of a PI from data provided by the financial YAHOO service, which presents historical data on the performance of FA on an investment exchange. This technique is described in detail in the following sections, which are organized as follows; sections 2 and 3 present the general definitions and methodology used respectively, and section 4 presents the results obtained.

2. MATERIALS AND METHODS

The developed application was developed entirely in MATLAB. This application takes a set of financial YAHOO data, which is processed with an optimization algorithm to find the most appropriate Investment Portfolio (IP) configuration. The following is a detailed description of the context for understanding the operation of the application.

2.1. Investment Portfolio

An IP is a set of documents representing Financial Assets (FA), in which an entity (person or company) invests an amount of money. One way of grouping this set of FA is done according to the risk profile of the investor. However, there are financial entities that already have their portfolios established and if the user decides to enter the entity adopts its level of risk. This level of risk sometimes indicates the time of return on investment and profits generated by the IP [1]. In addition, the level of risk allows investors to group is three profiles, which are:

Conservative: This group of people seeks a fixed and stable return over time, i.e. they do not run the risk that their IP will not achieve the proposed goal. They therefore benefit from IPs which have a minimum return and, in some cases, do not generate profits [2].

Moderate: This group is characterized by risking a little more than the conservatives, to obtain profits year after year that at least allow them to double their initial investment [2].

Aggressive: This group is made up of people who want to recover their initial investment quickly and make a profit. Typically, this type of investor seeks at least twice as much profit as a moderate profile person [2].

All IPs have a certain risk, since there is a possibility that the FA set will not generate profits. Therefore, each IP may have a different FA configuration, since the composition of the PI changes depending on the initial capital and profits expected by the investor. Considering that FAs are volatile, there are several steps for forming IP, which are described below [9].

- The first step is to evaluate the objectives and level of risk that the investor wants to assume. The objective is for the financial institution to filter the FA according to the initial characterization they have of the user, in order to carry out an appropriate preselection.
- The second step is to evaluate the FA considering their own characteristics, such as volatility, profitability, risks or prospects.
- In the third step it consists of selecting the FA according to the sector of preference for the financial entity or the investor, in order to avoid making investments in fields of action where there is not much experience and thus diminish the risk when creating the IP.
- The fourth step is to measure the profitability of the assets, which is usually done by reviewing their behavior in recent years through a historical.
- In the fifth and last step, the usefulness of the IP should be maximized by making a future hypothesis considering the probabilities of success and failure. Normally, the Markowitz projection is used, which indicates that diversifying the portfolio reduces risk.

Finally, one of the advantages of this work is that it allows the shareholder to jump to step four directly, i.e. from any FA set the application determines which is the appropriate configuration for the IP. This configuration is built automatically considering the historical behavior of the selected FAs from the YAHOO financial tool, this tool is free and is updated daily to show users the behavior of the investment exchange. The history is saved in plain text file format and pre-filtered to remove records with wrong characters. At the end of the pre-filtering stage, the IP with the highest profit margin is selected with an optimization algorithm, which is described in the following sections.

2.2. Optimization Algorithm

An optimization algorithm is a technique that allows you to find maximum or minimum values for a function or set of elements. In the financial sphere, this type of technique has a mathematical association, which allows the financial guild to predict the behavior of an IP from probabilistic assumptions, for example; Markowitz, LaGrange or Multi-criteria [9]. The aim of these techniques is to diversify the investment, i.e. to reduce the risk of IP without changing its profitability. One of the most common ways to perform these operations is based on the quadratic and parametric functions shown in Table 1 [9, 10].

As shown in Table 1, the set of values X are the FA to be optimized to build the IP, where σ is the variance that the utilities have had in a period , . The idea is to maximize the performance of the target function E considering the budget constraints (initial capital), budgetary constraints (FA value) and the non-negativity that indicates whether there is a margin of error for each X that is not defined in your domain. However, this type of strategies is limited, as they do not consider the fluctuations in real time that FAs can have and affect the performance of IPs [9].

Table 1. Parametric functions to optimize a function (based on [9]).

Objective function	*Parametric restrictions*	*Budgetary restrictions*	*Non-negative*
$\max E_p = \sum_{i=1}^{n} X_i E_i$	$\sigma^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} X_i X_j \sigma_{ij}$	$\sum_{i=1}^{n} X_i = 1$	$\forall X_i > 1, X_i = 1$

This has allowed the systematization of models or creation of computational strategies that optimize IP. On the one hand, optimization using mathematical models reduces the margin of error. However, they are stationary models that do not consider all parameters of actual behavior, which increases the risk of IP. On the other hand, computational strategies incorporating optimization algorithms are still under development. These include genetic algorithms that are bio-inspired techniques used to optimize problems even without knowing the type of function [4-8].

A conventional Genetic Algorithm (GA) is based on the Darwinian selection principle, where the strongest individuals are those who survive. In this case, the GA is a population algorithm, where everyone in the population is assigned an aptitude value, which identifies the individual's performance in attempting to solve a problem [11-15]. In addition, the GA is an iterative process where everyone undergoes a process of selection, crossing and mutation. The objective of selection is to take the set of individuals most suitable to solve the problem, the objective of crossing is to perform an exploitation of the function and mutation ensures that the population explores the function. The number of generations the algorithm is run is user-defined as shown in Algorithm 1.

Algorithm 1. Conventional genetic algorithm

P0 = Initial population;

CG = Number of Generations;

GA = Current generation;

P0←Stochastic values;

While GA<CG **do**

f←Evaluate(P0);

P1,f1←Selection(P0,f);

P1,f1←Crossover(P1,f1);

P1,f1←Mutation(P1,f1);

P0←P1;

GA++;

End While

There are variants of the GA that control the operators of mutation and crossing, in order to intensify the exploration and exploitation that is carried out in the functions to be optimized and thus to find global optimums. Some of these modifications are called multi-objective [12], multi-modal [13] or cultural [14] algorithms. In general, cultural algorithms are algorithms that operate in two spaces; the first is the space of the population where the characteristics of everyone in the population are defined and the second is the space of beliefs where the type of interaction between individuals in the population is established. In the same way as in GA, cultural algorithms are iterative algorithms that in each iteration of the algorithm try to find a better solution to the proposed problem. The functioning of the cultural algorithm is shown in detail in Algorithm 2.

Algorithm 2. Cultural algorithm (based on [14])

P0=Initial population;

ES=Define space of beliefs; CG=Quantity of Generations;

GA=Current generation;

CO=Number of operators;

P0←Stochastic values;

While GA<CG **do**

Communicate (P0, ES);

Adjust belief space (ES(CG)) f←Evaluate(P0);

P1,f1←Selection(P0,f);

P1,f1← Apply Crossover or Mutation operators (P1,f1,CO);

P0←P1;

GA++;

End While

This paper proposes a variant of the cultural algorithm where the space of beliefs is based on a random event that may occur during each iteration, by implementing several genetic operators. These features of the optimization algorithm are described in detail in the following section.

3. IMPLEMENTATION

As mentioned, the technique developed is based on a cultural algorithm that is used to maximize the performance of an IP, based on the pre-selection of a group of companies. In this case, the objective function is the utilities of the PI and the idea is to maximize performance by selecting the appropriate AF, i.e., an individual is composed of a random AF configuration and its aptitude value is the performance that those AFs can have over a period.

The FA are grouped by means of a stochastic selection made with a uniform distribution function and their suitability value is the average yield that the PI can get after six months of the initial investment. To estimate the behavior of the PI, an initial matrix is stored with the initial configuration of the portfolio, this variable is multiplied by itself and the result by the initial matrix. This operation is done 6 times where each time represents a month, is a process like the one done with the Markov chains. Next, a selection algorithm based on ranking is applied, this algorithm orders individuals from highest to lowest utility value and generates a new population from the stochastic selection of the ordered individuals.

Once the selection has been made, the algorithm proceeds to perform mutation and crossing operations. These operations are subject to a variable probability value, in other words, if the average proficiency value improves the probability of applying a crossover operator increases, otherwise the probability of applying a mutation operator increases. In total there are seven (7) mutation and crossing operators (three (3) crossing, four (4) mutation). When the algorithm enters the case of mutation or crossover operator selection, these can be selected randomly with the same probability rate. The proposed algorithm is executed several times determined by the user as shown in Algorithm 3 and a summary of its operation is presented in Fig.1.

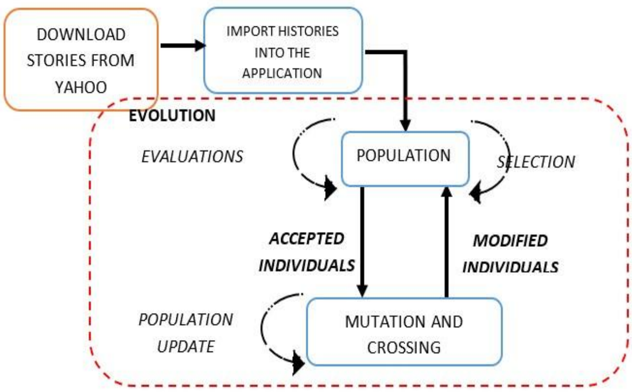


Fig. 1. Functioning of the developed algorithm

Algorithm 3. Developed algorithm

```
Evaluate function (Population)
    For i=1 to length (Population)
        T0=Population[i]
        For j=1 to 6
            T1=T0*T1
        End for
        F[i]=average yield(T1)
    End for
    Return (F)
End Evaluate

Optimizer Function (Historical)
    P0=Initial population;
    CG=Quantity of Generations;
    GA=Current generation;
    CO=Number of operators;
    GA←0;
    CO←7;
    P0←Create portfolios from the Historicals;
    While GA<CG do
        f←Evaluate(P0);
        P1,f1←Sort(P0,f);
        For i=1 up to length(P1)
            Threshold_0~U [0,1]
            V=V+f1[i]
            If V<Umbral_0 then
            P1,f1← Take from vector (P1[i],f1[i]);
        a←average (P1);
        If a<b then
            Threshold_1~U [0,1]
            If Threshold_1>0 & Threshold_1<=1/3 then
                P1,f1← Apply Crossover Operator 1(P1,f1);
            If Threshold_1>1/3 & Threshold_1<=2/3 then
                P1,f1← Apply Crossover Operator 2(P1,f1);
            If Threshold_1>2/3 & Threshold_1<=1 then
                P1,f1← Apply Crossover operator 3(P1,f1);
        Else
            Threshold_2~U [0,1]
            If Threshold_2>0 & Threshold_2<=1/4 then
                P1,f1← Apply mutation operator 1(P1,f1);
            If Threshold_2>1/4 & Threshold_2<=2/4 then
                P1,f1← Apply mutation operator 2(P1,f1);
            If Threshold_2>2/4 & Threshold_2<=3/4 then
                P1,f1← Apply mutation operator 3(P1,f1);
            If Threshold_2> 3/4 & Threshold_2<=1 then
                P1,f1← Apply mutation operator 4(P1,f1);
        b=a;
        P0←P1;
        GA++;
End Optimizer;

Main Function ()
    Download MATLAB GUI
    Start variables
    While the graphical interface is active do
        H=Enable the option to import Historical from a directory
        H=Filter(H)
        H=Normalize (H)
        Suitability=Execute Optimizer function (H)
        Graphing (Aptitude)
    End While
```

As observed the technique developed has several operators of crossing and mutation, these are described in detail in Table 2. Finally, the results obtained were compared with the techniques of optimization of hill ascent and simulated tempering. On the one hand, the choline ascent technique is an iterative method that compares two individuals, stores the better of the two, replaces the discarded individual with another randomly selected individual, and repeats the procedure. On the other hand, the simulated tempering technique has a function like that of ascending a hill, but the selection of the individual depends on a probability value that decreases exponentially over time [4]. The results obtained in each experiment are described in detail in the following section.

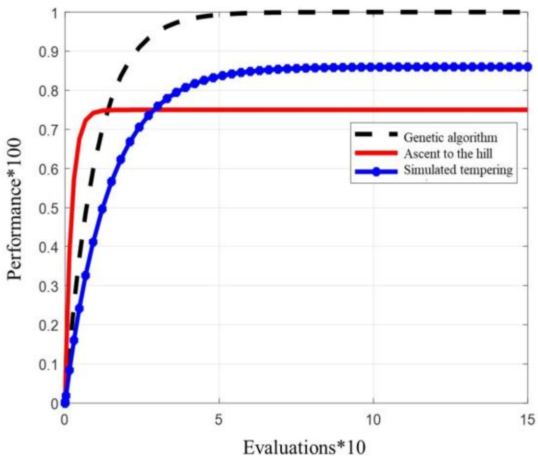
***Table 2**. Description of the mutation and crossing operators.*

Operator	**How it works**
	If $A = [A_1\ A_2\ A_3 \ldots A_n]$ and $B = [B_1\ B_2\ B_3 \ldots B_n]$ then
Crossing 1	A point is selected randomly and the components of the individual vectors are exchanged, so if the point is equal to 1 then $A = [A_1\ B_2\ B_3 \ldots B_n]$ and $B = [B_1\ A_2\ A_3\ \ldots A_n]$
Crossing 2	Two points are randomly selected and components of individual vectors are exchanged, so if the points are 1 and 3 you have to $A = [A_1\ B_2\ A_3 \ldots A_n]$ and $B = [B_1\ A_2\ B_3\ \ldots\ B_n]$
Crossing 3	Invert the components of the vectors and cross them in a point, So if the point is equal to 1 you have to $A = [B_n \ldots\ B_3\ B_2\ A_1]$ y $B = [A_n \ldots\ A_3\ A_2\ B_1]$
	If $A = [A_1\ A_2\ A_3 \ldots A_n]$ in that case
Mutation 1	Randomly select a component and change it to another one, so if the component is equal to 1 you have to $A = [A_1\ C_x\ A_3 \ldots A_n]$
Mutation 2	Change one individual for another $A = B$
Mutation 3	Invert the components of an individual $A = [An \ldots.\ A_3 A_2 A_1]$
Mutation 4	Do not modify the individual.
Note: Note that each individual component is a financial asset, therefore, there are no modifications to the components of each asset.	

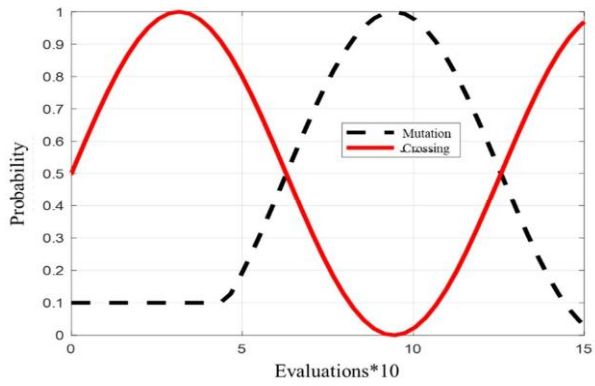
4. RESULTS AND DISCUSSION

Initially the performance of each optimization technique was measured by trying to find an appropriate configuration of each IP, this result is presented in Fig. 2a. This graph shows the result of 150 evaluations of the aptitude value and the probability that the IP utilities are maximized, Fig. 2b. Shown the behavior of mutation and crossing operators during the execution of the proposed technique, which is composed of populations of ten (10) individuals.

As described in the previous section, individuals are generated randomly, i.e. the configurations of each IP change according to the utilities measured from the historical at different time intervals. The conformation of the portfolio for only two individuals of the initial population is shown in Fig. 3.



(a) Behavior of the suitability value.



(b) Behavior of genetic operators.

Fig. 2. Behavior of optimization techniques.

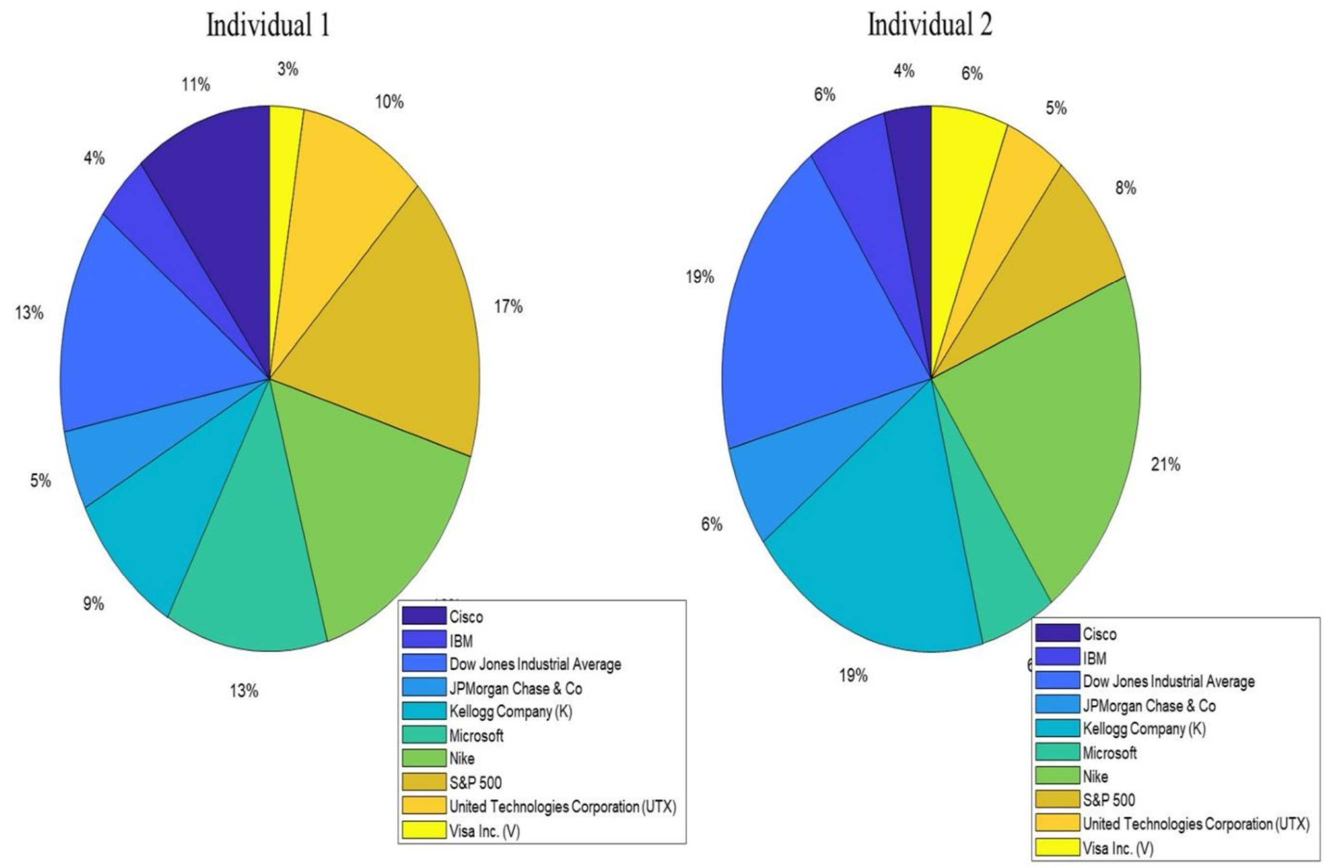


Fig. 3. Configuration of individuals initially.

Finally, the best portfolio configuration found by the three (Cisco, IBM, Dow Jones Industrial Average, JPMorgan Chase optimization techniques in Figure 4a, 4b and 4c is shown at the & Co, Kellogg Company (K), Microsoft, Nike, S&P 500, end of the 150 proposed executions. Each experiment United Technologies Corporation (UTX), Visa Inc. (V)) during compared and evaluated the performance of ten companies the last two years.

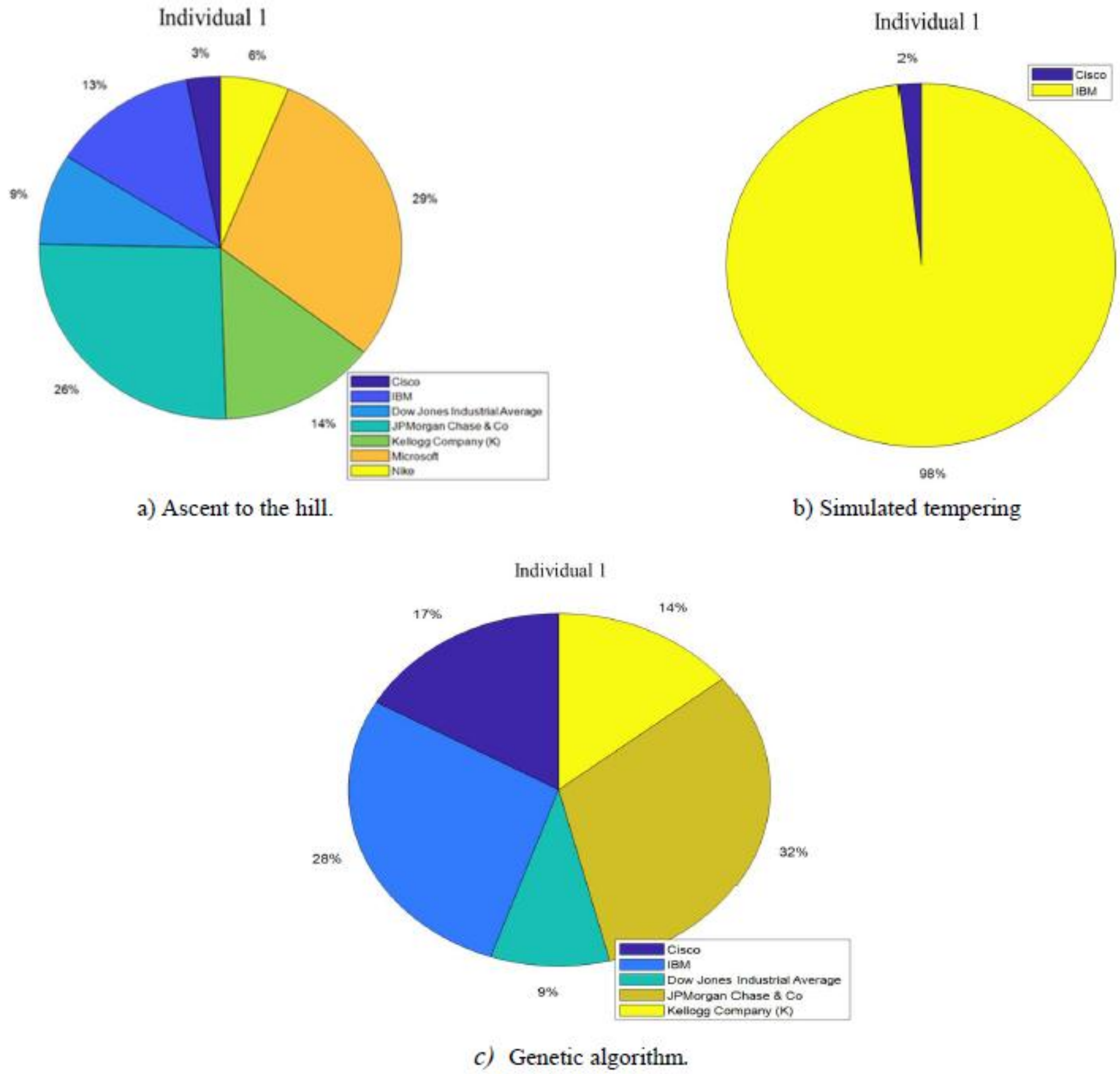


Fig. 4. Best individual found by each technique.

5. CONCLUSIONS

As presented in Fig. 2a. the proposed technique finds configurations of financial assets that allow the user to build an investment portfolio with a 100% return. This percentage indicates that the user has the possibility of doubling his profits in a period of six months, by investing in the best portfolio found during the execution of the proposed technique. Bearing in mind that the application can evaluate up to ten different companies, i.e. that the restriction of the type of sector to invest in one is subject to evaluation by the user.

As can be seen in Fig. 2b. the proposed technique has fluctuations between the number of crosses and mutations, i.e. when the probability of crossing increases the probability of mutation decreases. This phenomenon occurs because the algorithm performs exploitation of certain areas of the function to optimize and reaches the point where it can no longer optimize, so it performs exploration actions to see if it finds new areas within the function or configuration of individuals that can optimize. In addition, this phenomenon is advantageous for the proposed technique, since, as shown in Fig. 2a. the algorithm of hill climb and simulated tempering converge rapidly, because they cannot perform exploration operations to find new optimization zones within the function to be optimized.

Finally, as shown in Fig. 4 all techniques find portfolio configurations in which to invest. However, the performance of the aptitude function must be considered when choosing the portfolio that has the most utilities.

REFERENCES

[1] P. U. Shinde and S. R. Deshmukh, "Risk management in electricity market by portfolio optimization," 2014 Annual IEEE India Conference (INDICON), Pune, 2014, pp. 1-6. doi: 10.1109/INDICON.2014.7030542

[2] K. Ye, H. Li, X. Shi and N. Shi, "Resource Allocation Problem in Port Project Portfolio Management," 2014 Seventh International Joint Conference on Computational Sciences and Optimization, Beijing, 2014, pp. 159-162. doi: 10.1109/CSO.2014.36

[3] R. Fang, L. Nianping and W. Xiaoqing, "Numerical Simulation of Natural Ventilation of a Folio Multi-Sash Mid-Pivoted Window with Vertical Deflectors in a Room," 2011 Third International Conference on Measuring Technology and Mechatronics Automation, Shangshai, 2011, pp. 717-720. doi: 10.1109/ICMTMA.2011.463

[4] Bharti and D. K. Sharma, "Searching boolean function using simulated annealing and hill climbing optimization techniques," 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, 2016, pp. 62-64. doi: 10.1109/ICACCCT.2016.7831601

[5] T. Cui, S. Cheng and R. Bai, "A combinatorial algorithm for the cardinality constrained portfolio optimization problem," 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, 2014, pp. 491-498. doi: 10.1109/CEC.2014.6900357

[6] J. Chu-Xin, C. Wan-Yi and Y. Shu-Jing, "Robust Portfolio Selection Based on Optimization Methods," 2018 37th Chinese Control Conference (CCC), Wuhan, 2018, pp. 8359-8364. doi: 10.23919/ChiCC.2018.8483072.

[7] Y. Zaychenko and I. Sidoruk, "Problem of fuzzy portfolio optimization and its solution with application of forecasting methods", 2015 Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies (CSIT), Lviv, 2015, pp.97-103. doi: 10.1109/STC-CSIT.2015.7325442

[8] G. A. V. Pai, "Multi-objective Differential Evolution Based Optimization of Risk Budgeted Global Asset Allocation Portfolios," 2014 2nd International Symposium on Computational and Business Intelligence, New Delhi, 2014, pp. 17-20. doi: 10.1109/ISCBI.2014.11

[9] H. Kamili and M. E. Riffi, "A comparative study on portfolio optimization problem," 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-8. doi: 10.1109/ICEMIS.2016.7745339

[10] J. J. Liang and B. Y. Qu, "Large-scale portfolio optimization using multiobjective dynamic mutliswarm particle swarm optimizer," 2013 IEEE Symposium on Swarm Intelligence (SIS), Singapore, 2013, pp. 1-6. doi: 10.1109/SIS.2013.6615152.

[11] S. Kapil, M. Chawla and M. D. Ansari, "On K-means data clustering algorithm with genetic algorithm," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, 2016, pp. 202-206. doi: 10.1109/PDGC.2016.7913145

[12] M. Dario Arango Serna and J. Andres Zapata Cortes, "Multiobjective Model For The Simultaneous Optimization Of Transportation Costs, Inventory Costs And Service Level In Goods Distribution," in IEEE Latin America Transactions, vol. 15, no. 1, pp. 129-136, Jan. 2017. doi: 10.1109/TLA.2017.7827916

[13] E. Dilettoso and N. Salerno, "A self-adaptive niching genetic algorithm for multimodal optimization of electromagnetic devices," in IEEE Transactions on Magnetics, vol. 42, no. 4, pp. 1203-1206, April 2006. doi: 10.1109/TMAG.2006.871672

[14] S. Upadhyayula and Z. Kobti, "Population Migration Using Dominance in Multi-population Cultural Algorithms," 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, 2015, pp.614-617. doi: 10.1109/ICMLA.2015.102

[15] X. Shao, Z. Ren and R. Wang, "New control allocation algorithm for three-dimension problem based on genetic algorithm," Proceedings of the 33rd Chinese Control Conference, Nanjing, 2014, pp. 3766-3770. doi: 10.1109/ChiCC.2014.6895566