European Research Studies,
Volume XVII, Issue (4), 2014 *pp. 55-66*

Clustering Stock Exchange data by Using Evolutionary Algorithms for Portfolio Management

Malek Khojasteh Nejad [1]

Abstract:

In present paper, imperialist competitive algorithm and ant colony algorithm and particle swarm optimization algorithm have been used to cluster stocks of Tehran stock exchange. Also results of the three algorithms have been compared with three famous clustering models so called k-means, Fcm and Som. After clustering, a portfolio has been made by choosing some stocks from each cluster and using NSGA-II algorithm. Results show superiority of ant colony algorithms and particle swarm optimization algorithm and imperialist competitive to other three methods for clustering stocks. Due to diversification of the portfolio, portfolio risk will be reduced while using data chosen from the clusters. The more efficient the clustering, the lower the risk is. Also, using clustering for portfolio management reduces time of portfolio selection.

Key Words: *Portfolio Management, Data mining, Imperialist Competitive Algorithm, Ant Colony Algorithm, Particle Swarm Optimization Algorithm*

JEL Classification : *F21, G11*

[1] MA in economics, University of Sistan & Baluchestan, Zahedan, Iran, Email: malek.kh16@gmail.com

1. Introduction

Nowadays, choosing a suitable portfolio is one of the most important issues that investors of financial markets face with it. Markowitz [1] was the first person who outlined diversification in portfolio. He believed that investors pay attention simultaneously into risk and revenue. Investors are looking for increasing the expected revenue and reducing the risk. In Markowitz Model, the mean is a standard for the revenue; standard deviation and variance are standards for measuring risk.

After Markowitz model, people such as Sharpe [2], Elton et.al [3] and Konno [4] offered new solutions to solve problems of Markowitz model and portfolio selection. Among methods used recently in choosing portfolio are single- and multi-objective evolutionary algorithms. One of advantages of these algorithms is their non-linearity. These algorithms are very efficient for choosing portfolio when there are a number of assets.

In methods used for solving optimization problems, experiences of nature and their Imitation has been applied. In order to solve problems in evolutionary algorithms, developmental process of animals, plants and organisms generally has been inspired from the nature. Since organisms have developed their own solutions for solving problems during thousands of years, they have obtained a relatively optimal solution for their lives. As one of methods for solving problems inspired from natural development, evolutionary algorithms are able to find an optimal response and solve complex and time-consuming calculations. Traditional methods cannot deal with it. Inheritance and reproduction, random change and natural selection are operators that make possible transition from one generation of organism to another. In fact, chance and natural selection are considered as two important factors in the evolution. [5]
The present research is aimed to use clustering to improve performance of NSGA-II algorithm in choosing portfolio. Clustering analysis is a method for grouping data or observations regarding their similarity or proximity. By clustering analysis, data or observations are divided into homogenous and heterogeneous groups. By clustering, we are going to find similar data in order to identify behaviours well and get a better result. In present research, stock exchange has been clustered using imperialist competitive algorithm, ant colony algorithm, particle swarm optimization algorithm and k-means, Fcm, Som methods. Then, due to similarity of stocks in each cluster, few stocks have been chosen from each cluster. Therefore, among chosen stocks, the portfolio has been selected using NSGA-II algorithm.

Results indicate that in practice, clustering reduces the time required for choosing portfolio. Therefore the risk will be reduced due to diversification of portfolio.

2. Literature Review

In this section, studies done on portfolio management and clustering methods will be dealt with in brief.

2.1. Portfolio Management

By a quantitative definition of investment risk for investors and selection of assets and portfolio management, Markowitz offered a mathematical approach. According to him, investors can obtain an efficient portfolio per certain revenue by minimizing portfolio risk or per a certain risk by maximizing portfolio revenue [1]. Many studies have been done on portfolio management using evolutionary algorithms. Chiam .et.al [6] have offered a new method for portfolio management using NSGA-II algorithm. In this method, the number of stocks present in each portfolio can be chosen. Duran .et.al [7] and Chang .et.al [8] indicated efficiency of evolutionary algorithms for portfolio management when many stocks are under study. Sadeghi and Zandieh [9] offered a game theory model in order to manage portfolio in product market. Östermark [10] have used fuzzy model to manage portfolio

2.2. Clustering Techniques

K-means method is one of the famous methods of clustering. This method was offered by McQueen [12] and it has been used independently or in combination with other methods in different sciences. Krishna and Morty [13] presented GKA model by combining genetic algorithm and k-means method. Its difference with clustering of genetic algorithm is that mean data in each cluster has been used as the new centre of cluster rather than random selection of cluster centres. Kananga .et.al [14] presented another effective method to improve structure of K-means. By using a method called filtering algorithm, they reached a better and more rapid way for optimal number of clusters in multidimensional space.

Fcm is another method of clustering. After definition of Fuzzy set by Zadeh[15], the first step was taken by Ruspini [16] in order to use fuzzy systems for clustering. The aim is to analyze clusters based on standard of error least square in Isodata algorithm. Then, Bezdek and Joseph [17] completed Fcm algorithm.

Evolutionary algorithms are new and efficient tools for clustering. Some efforts were done by Krovi [18] for clustering by evolutionary algorithm; he used different methods for improving and applying genetic algorithm in clustering. Maulik and Bandyopadhyay [19] offered a genetic clustering method for automatic evaluation of clusters (choosing optimal number of clusters during implementing the program). Garai and Chaudhuri [20] present a two-step method. In the first step, main data

were divided into certain groups. In the second step, separated data were converted into a clustering combination of K cluster. Different algorithms have been used in different studies. Kalyani, and Swarup [21] and Cura [22] applied Particle swarm optimization algorithm for clustering. Shamshirband .et.al [23] used imperialist competitive algorithm and Ye and Mohamadian [24] and Ji .et.al [25] have used ant colony algorithms for data clustering/

3. Methodology

3.1 K-means

K-means algorithm allocates a point to the cluster which centre is closer. In K-means algorithm, k member (k is number of clusters) is chosen randomly among n members as cluster centres. Then remaining N-K members allocate to the closest cluster. After allocation of all members, cluster centres are recalculated and members will be allocated based on new centres in clusters. It continues until cluster centres are stable. The main advantage of this algorithm is its easiness and speed so that it can be applied for large data groups. The disadvantage of this algorithm is that the algorithm obtains different results in each implementation.

3.2 Fuzzy C-means

Fuzzy clustering and Fcm clustering methods have many applications in different problems of clustering. The aim of this algorithm is to separate data of $\{X_1, \ldots . X_n\} \subset R^s$ into C cluster number based on minimization of the least interval function as follows:

(1) $$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{m} \parallel X_k - V_i \parallel^p$$

Where m (its amount is larger than one) is fuzziness parameter. Also $V_i \in R^s$ is in cluster centre. $\mu_{ik} \epsilon [0,1]$ is the number of data per cluster and p is power of Euclidean distance.
Optimal U and V values are obtained using optimization algorithm. The most efficient method of fuzzy clustering for optimizing the equation 1 is Fcm method. In order to optimize function of $J_m(U,V)$ on parameters U and V, optimization algorithm is done in two steps for estimation of U and V. cluster centres (v) in its stage concerning value of U in (r-1) stage is calculated as follows:

$$V_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_k}{\sum_{k=1}^{n} (\mu_{ik})^m} \quad (2)$$

Then, new value of U (using calculated value for V in previous part) is obtained by following relation:

$$\mu_{ik} = \sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{-2/(m-1)} \quad (3)$$

3.3 Self Organizing Maps (SOM)

SOM is a non-controlled nervous network that consists of nervous cells in a regular grid with low dimensions. Each neuron has the n-dimensional weight vector which is n-fold of dimensions of input vector. Weight vectors connect the input layer to the output layer. Neurons are connected to each other by a neighbourhood function. Each input vector is activated based on the highest neuronal similarity in output layer which is called the winner cell. The similarity is measured by Euclidean distance between two vectors:

$$D_j = \sum_{i=1}^{n} \| W_{i,j} - X_i \|^2 \quad (4)$$

X_i is the ith input vector, $W_{i,j}$ is the weight vector that connects i input into output neuron j and D_j is the sum of Euclidean distance between X_i input and weight vector connected to jth output cell that is called map unit.

3.4 Ant Colony Algorithm

Ant Colony Optimization (ACO) was offered by Dorigo .et.al [26] in order to solve traveling salesman problem. This algorithm has been inspired from finding food by real ants so that each agent (factor) is an artificial ant. Trials of biologists on Argentinean ants showed that if two ways are considered for ants from their formicary to food source, most of ants will choose the shorter way after some while (within few minutes).

The number of ants will increase by the increasing difference of two ways. Because ants release a chemical called Pheromone when coming and going the formicary, this chemical is evaporative. In other words, when ants reach the decision making point, they will choose the rout randomly because there is no pheromone on the routs. But after a while, the same ants release pheromone based on shorter rout and better food source. At this time, selection is made by probability meaning that the rout with higher pheromone will be chosen by most of ants. This behaviour is has a simulative effect because repetitive selection of a rout will increase probability of choosing that rout again.

Therefore, more pheromone released on the shorter rout will be stored and the shorter rout is more attractive than the farther one. Dorigo used this simple idea to

find good solutions for difficult optimization problems and offered ant system algorithm as the first version of ACO algorithm. In this algorithm, the task of each artificial ant (like the real one) is to find the shortest route between a pair of nods in a graph in which the problem has been plotted appropriately .This algorithm is discrete that has been designed for solving discrete optimization problems. Socha and Dorigo [27] introduced continuous version of ant algorithm so called ACOR ACOR attempts to follow meta-heuristic ACO. Its structure allows users to solve problems of discrete-continuous optimization. The main idea of ACOR is to use continuous probability distribution using probability density function instead of discrete probability distribution.

3.4 Imperialist Competitive Algorithm

Imperialist competitive algorithm (ICA) is one of the newest smart optimization algorithms that have been introduced by Atashpaz-Gargari and Lucas [28] in evolutional calculations and computational intelligence. The main idea of this algorithm is to simulate political process of imperialism. Like genetic algorithm that simulates biological evolution, political development has been used in imperialist competitive algorithm. This algorithm begins with a number of random populations so called the country.

The best elements of the population are chosen as imperialists. The remaining is considered as colonies. Imperialists depending on their power, attract colonies with a certain process. Total power of any empire depends on country of imperialist and its colonies. By formation of primary empires, the competition between imperialists is started. Any empire that cannot be successful in the competition and increase its power will be removed from the competition. Therefore, survival of an empire depends on its power for attracting colonies of other empires and controlling them. As a result, in imperialist competition, power of great empires is increased gradually and weak empires will be removed. In order to increase their powers, the empires have to progress their colonies. When colonies are travelling towards the country of the imperialist, they may reach a position better than the imperialist.

In this case, country of imperialist is displaced by country of colony and the algorithm continues with imperialist country in new situation. Finally, all empires will be failed and only one empire is available and other countries will be under control of this empire and a convergence will be brought about.

3.5 Particle Swarm Optimization Algorithm

Particle swarm optimization algorithm is an evolutional algorithm to optimize nonlinear functions and it has been offered based on social behaviour of birds. This

algorithm was offered by Kennedy and Eberhart [29] that derived from behaviour of particle swarms such as bird flocks. Therefore, in a flock, one bird (leader) has the best position and other birds try to approach the leader and make their position better.
If one bird can reach a position better than the leader, it can be chosen as the leader. In this algorithm, there are some things that are called particles and they distributed in search space of the function that we want to minimize or maximize its amount. Each particle calculates amount of target function in the space where it has been placed. Then it chooses a direction for its movement using information of current place and the best place where it has been in the past and also information of one or more particles (the best particles). All particles choose a direction for movement. After that, one stage of algorithm is finished. These stages are repeated several times in order to reach the appropriate solution. In fact, particle swarm that searches minimum amount of a function acts as flocks that are searching foods.

3.6 Genetic algorithm

Genetic algorithm is simulation of biological evolution derived from Darwin natural evolution. In this method, among available responses, good ones that improve target function are chosen for next generation of responses and this evolutionary process is repeated to find the optimal response. This algorithm was firstly offered by Holland [30].

Genetic algorithm has been derived from inheritance of parent characters to children by combining parents' chromosomes. In each generation, chromosomes are evaluated by some standards of goodness of fit. In order to make the next generation, new chromosomes that are called offspring are formed either by crossover operator from chromosome of the current generation or by correction of a chromosome using mutation operator. After several generations, algorithm converges towards the best chromosome that presents an optimal solution for the problem. Due to certain problems in classic techniques of optimization of multi-objective evolutionary algorithms, algorithms have been suggested that are able to implement complex problems in multi criteria spaces. Non-dominated or genetic algorithms are among those for multi-objective optimization. After presenting the first version of this algorithm in 1995, developers of this algorithm among whom Deb is more famous than others presented the second version of the algorithm abbreviated by NSGA-II .

This algorithm has been converted into a multi-objective algorithm by adding necessary operators to single purpose genetic algorithm. Instead of finding the best answer, it offers a group of answers that are known as Pareto-Front. These two operators are the operator that allocates a better sort to the members of population

based on non-dominated sorting and the operator that keeps various answers with equal grade.

4. Data Description and Experimental Results

In present paper, the companies under study were chosen from those participated in Tehran stock exchanges. Those companies participated in stock exchange from May 2013 to September 2014 were chosen for this research. It has been attempted to choose companies from all industries and groups participated in Tehran stock exchange. As a result, 200 companies were chosen. Variables used for clustering the stocks are asset revenue, net profit margin, profit per stock, growth rate of gross benefit, growth rate of profit per stock, ratio of market value to the book value, ratio of the price to the return, revenue of shareholders' equity.

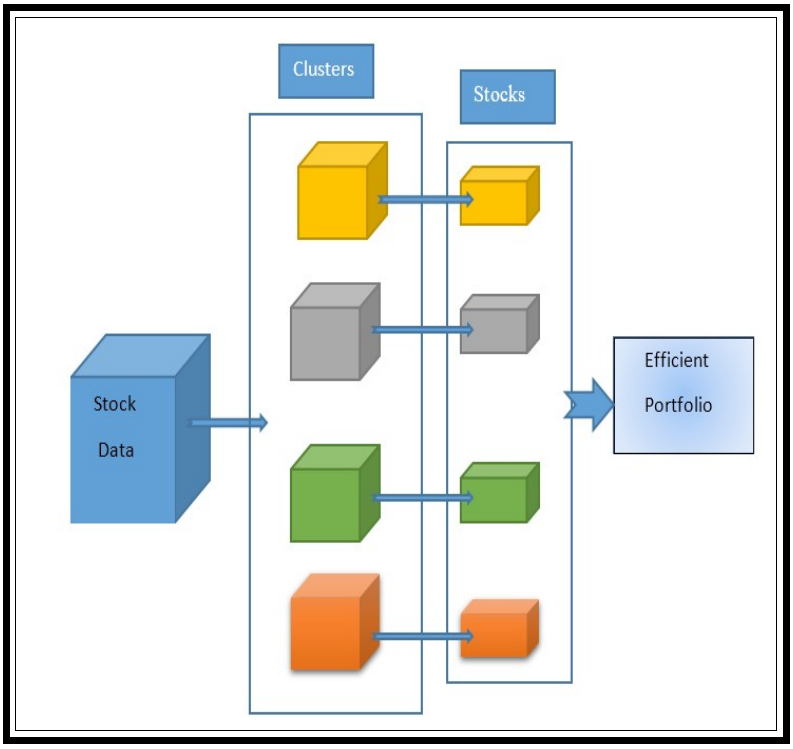
Three algorithms of imperialist competition, Particle swarm optimization algorithm, and ant colony have been used for clustering. 8 clusters have been considered. Results of clustering using these algorithms have been compared with results of three famous clustering models k-means, Fcm, Som in table 1.

Table 1. Results of Clustering Models

Model	Fcm	Som	K_means	Pso	Ant koloni	Ica
Davies Bouldin	0.4706	0.4738	0.4756	0.4192	0.4078	0.4223

Considering the amount obtained by Davies Bouldin, the best results were obtained by ant colony algorithm, Particle swarm optimization algorithm, and imperialist competitive algorithm, FCM, Som and K-means respectively. These results indicate that evolutionary algorithms are more efficient than other methods for stock clustering. Some stocks were chosen from each cluster for every clustering method. A portfolio containing 10 stocks and their weights has been determined using NSGA-II algorithm. In fig 1, this process has been shown briefly.

Figure 1. Creation of Efficient Portfolios



In order to study the effect of clustering on the risk of portfolio, cost function has been designed in a way that the total return is equal to 2.13e-05 (meaning that number of ten stocks and weights has been chosen such that sum of return of the portfolio is equivalent to 2.13e-05 so that the risk is reduced). 2.13e-05 value has been chosen randomly. The cost function has been indicated in equation 5.

$$\text{Min } F_i = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}$$
$$\text{Max } F_2 = \sum_{i=1}^{N} w_i \mu_i \quad (5)$$

Subjected to

$$\sum_{i=1}^{N} w_i = 1$$
$$0 \leq w_i \leq 1 \quad i = 1, \ldots, N$$

Where N is the number of stocks available, μ_i is the expected return of stock i, σ_{ij} is the covariance between stocks i and j and w_i is the decision variable denoting the composition of the portfolio.

Amount of risk of the selected portfolio has been indicated in table 2 by NSGA-II algorithm for different methods of clustering. Also, among all stocks, the selected portfolio risk has been presented in this table (without using clustering).

Table 2. Risk of the Selected Portfolio

Model	Fcm	Som	K_means	Pso	Ant koloni	Ica
Total Risk	3.10e-05	3.57e-05	4.22e-05	2.16e-05	1.42e-05	2.85e-05

As seen in results, using clustering for portfolio diversification reduces portfolio risk. Also results indicate that the more precise the clustering, the lower the risk. Empirical results show time reduction that used clustering for choosing portfolio.

5. Conclusion

One of methods recently used for selection of portfolio is NSGA-II algorithm. In present research, it has been tried to improve efficiency of the model using clustering. Stocks are divided into some clusters using clustering. The stocks present in each cluster are very similar to each other and are very different with stocks in other groups. Considering that the stocks in each group are similar to each other, a number of stocks are chosen from each cluster in order to form portfolio that have characteristics of all stocks present in that cluster, among chosen stocks, the appropriate portfolio is selected using NSGA-II algorithm. Time of choosing portfolio will be reduced due to reduction of number of stocks. Portfolio diversification reduces portfolio risk. Three algorithms of imperialist competition, Particle swarm optimization, ant colony have been used for clustering. Results of these algorithms have been compared to those of three methods k-means, Fcm, Som. Clustering results indicate superiority of algorithms of ant colony, Particle swarm optimization, imperialist competition, FCM, Som and K-means algorithms.

References

Atashpaz-Gargari E., Lucas C. (2007) 'Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition', *Evolutionary Computation*.

Baltas, N. K., Silipo, D., Kapetanios, G., and Leonida, L., (2014) "Assessing Bank Efficiency and Stability", *6th Conference of the International Finance and Banking Society (IFABS)*, *Alternative Futures for Global Banking: Competition, Regulation and Reform, University of Surrey, Guildford, UK*.

Bezdek J. C., Dunn J. C. (1975), "Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions", *Computers*, 100(8), 835-838.

Chang T. J., Yang S. C., and Chang K. J. (2009), 'Portfolio Optimization Problems in Different Risk Measures Using Genetic Algorithm', *Expert Systems with Applications*, 36(7), 10529-10537.

Chiam S. C., Mamun A. A., and Low Y. L., (2007) 'A Realistic Approach to Evolutionary Multi-Objective Portfolio Optimization', *Evolutionary Computation*

Cura, T. (2012) 'A Particle Swarms Optimization Approach to Clustering', *Expert Systems with Applications* 39(1), 1582-1588.

Dorigo M., Maniezzo V. and Colorni A. (1996) 'Ant System: Optimization by a Colony of Cooperating Agents', *Systems, Man, and Cybernetics*, Part B: Cybernetics, 26(1), 29-41.

Duran F. C., Cotta C. and Fernández A. J. (2009) 'Evolutionary Optimization for Multi-Objective Portfolio Selection under Markowitz's Model with Application to the Caracas Stock Exchange', 489-509.

Eiben A. E., Smith J. E. (2003) 'Introduction to Evolutionary Computing'

Elton E.J., Gruber M.J., and Padberg M. (1976) 'Simple Rules for Optimal Portfolio Selection', *Journal of Finance*, 31, 1341–1357.

Garai G., Chaudhuri B. B. (2004) 'A Novel Genetic Algorithm for Automatic Clustering', *Pattern Recognition Letters*, 25(2), 173-187.

Holland J. H. (1974) "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence', *University of Michigan Press*.

Ji J., Song X., Liu C. and Zhang X. (2013) 'Ant Colony Clustering with Fitness Perception and Pheromone Diffusion for Community Detection in Complex Networks', *Physica A: Statistical Mechanics and its Applications,* 392(15), 3260-3272.

Kalyani S., Swarup K. S. (2011) 'Particle Swarm Optimization based K-means Clustering Approach for Security Assessment in Power Systems', *Expert Systems with Applications*, 38(9), 10839-10846.

Kanungo T., Mount D. M., Netanyahu N. S., Piatko C. D., Silverman R. and Wu A. Y. (2002) 'An Efficient K-means Clustering Algorithm: Analysis and Implementation., Pattern Analysis and Machine Intelligence', *IEEE Transactions*, 24(7), 881-892.

Kennedy J., Eberhart R.C, (1995) 'Particle Swarm Optimization,' *Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia

Konno H. (1990) 'Piecewise Linear Risk Function and Portfolio Optimization', *Journal of Operational Research Society Japan*, 33(2), 139–156.

Krishna K., Murty M. N. (1999) 'Genetic K-means Algorithm, Systems, Man, and Cybernetics, Part B: Cybernetics', *IEEE Transactions*, 29(3), 433-439.

Krovi R. (1992) 'Genetic Algorithms for Clustering: a Preliminary Investigation.' System Sciences', *Proceedings of the Twenty-Fifth Hawaii International Conference*

MacQueen J. (1967) 'Some Methods for Classification and Analysis of Multivariate Observations', Proceedings *of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14).

Maulik, U., Bandyopadhyay S. (2000) 'Genetic Algorithm-based Clustering Technique', *Pattern recognition*, 33(9), 1455-1465.

Östermark, R. (1996) "A Fuzzy Control Model (FCM) for Dynamic Portfolio Management', *Fuzzy sets and Systems*, 78(3), 243-254.

Ruspini E. H. (1969) 'A New Approach to Clustering', *Information and control* 15(1), 22-32.

Sadeghi, A., Zandieh, M. (2011) 'A Game Theory-based Model for Product Portfolio Management in a Competitive Market', *Expert Systems with Applications*, 38(7), 7919-7923.

Shamshirband, S., Gocić, M., Petković D., Javidnia H., Ab Hamid S. H., Mansor Z. and Qasem S.N. (2015) 'Clustering Project Management for Drought Regions Determination: A Case Study in Serbia', *Agricultural and Forest Meteorology*, 200, 57-65.

Sharpe, W.F. (1996) 'Mutual Fund Performance', *Journal of Business*, 39(1), 119-138.

Socha, K, Dorigo, M. (2008) 'Ant Colony Optimization for Continuous Domains', *European Journal of Operational Research*, 185 (3), 1155-1173.

Thalassinos, I.E., Pociovalisteanu, D.M., (2007) "A Time Series Model for the Romanian Stock Market", *European Research Studies Journal Vol. X,* Issue 3-4, 57-72.

Thalassinos, I.E., Politis, D.E., (2011) "International Stock Markets: A Co-integration Analysis," *European Research Studies Journal*, Vol. XIV, (4), 113-130.

Thalassinos, I.E., Venediktova, B., Staneva-Petkova, D., and Zampeta, V. (2013) 'Way of Banking Development Abroad: Branches or Subsidiaries', *International Journal of Economics and Business Administration*, 1(3), 69-78.

Ye Z, Mohamadian H. (2014) 'Adaptive Clustering based Dynamic Routing of Wireless Sensor Networks via Generalized Ant Colony Optimization'. *IERI Procedia* 10, 2-10.