Journal of Theoretical and Applied Information Technology
10th July 2015. Vol.77. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MODIFICATION OF NEURAL NETWORK ALGORITHM USING CONJUGATE GRADIENT WITH ADDITION OF WEIGHT INITIALIZATION

[1]AZWAR RIZA HABIBI, [2]TRISILOWATI, [3]RATNO BAGUS EDY WIBOWO
[1]P.G. Student, Department of Mathematics, Brawijaya University, Indonesia
[2,3]Lecturer, Department of Mathematics, Brawijaya University, Indonesia
E-mail: [1]riza.bj@gmail.com, [2] trisilowati@ub.ac.id, [3]rbagus@ub.ac.id

ABSTRACT

This paper develops neural network (NN) method using conjugate gradient (CG) with combination of particle swarm optimization (PSO) and genetic algorithm (GA). The combination of PSO and GA is used for weight initialization to improve the computational process and minimize the errors. CG method can change every learning rate of neural network, so that the addition of CG can increase the rate of convergence. PSO, by its calculation velocity, can get quickly the solutions and GA act to expand the searching area of PSO solution. This new algorithm is called CN-PSOGA. There are some criterions used to compare the CN-PSOGA algorithms with others, i.e. accuracy degree, number of iterations, and computation time produced by each algorithm. Simulation results show that the proposed algorithm can increase the accuracy of solution approach.

Keywords: *Neural Network, Conjugate Gradient, Particle Swarm Optimization, Genetic Algorithms.*

1. INTRODUCTION

NN is an adaptive computation of human brain behavioural system which is used to develop the new model of algorithm to find the optimal solutions. Many papers use NN algorithm as the combination of some algorithms which is used to help finding the best solution. Nawi et al. [1] modified NN algorithm with CG algorithms. There are several methods which can be used to determine the direction of gradient for CG algorithm, for example, conjugate gradient Fletcher Reeves (CGFR). The modification of CGFR algorithm at the initial search direction by adding the value is called adaptive gain on activation functions. Al-Bayati et al. [2] modified CG algorithm based on Nawi et al., which is used for a neural network-backpropagation algorithm. A simple modification CG on NN can substantially improve the training efficiency. This also shows that the direction of local search which is modified by an adaptive gain with proper activation function, can significantly improve the convergence rate.

PSO algorithm mimics the behaviour of animal living in groups. Commonly, PSO algorithm is simple to implement, effective and very fast to converge [3],[4], however its drawback is that the algorithm may prematurely converge. GA is an optimization method based on the natural biological evolution mechanism. This algorithm provide a robust search procedure and does not depend on specific areas [5], however it has less ability to control the convergence when compared to PSO [4]. In 2011, Singh et al. [6], introduced the binary particles for PSO optimization with crossover to solve discrete functions. Five types of binary crossover operator are used to binary optimization PSO to investigate the working of hybrid algorithm at benchmark functions. The result shows that the hybrid algorithm provides a better solution to some benchmark functions.

This paper develops the NN algorithm that uses CG, comprising merits of PSO and GA as weight initialization, and this algorithm is called the conjugate gradient-neural network-PSO-genetic algorithm (CN-PSOGA). Furthermore, the data of the number of Juanda flight passengers (DATA 1) and the amount of rainfall (DATA 2) are used for CN-PSOGA simulation. It's to see the superiority of CN-PSOGA algorithm in getting solutions and having the minimum error, and then the CN-PSOGA will be compared with other algorithms.

The rest of the paper is organized as follows. In the next section, material and method are discussed. While in the third section, CN-PSOGA algorithm is

Journal of Theoretical and Applied Information Technology
10th July 2015. Vol.77. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

described. Result is presented in section four followed by conclusion in the final section.

2. MATERIALS AND METHOD

Data used for training set is obtained from BPS (Agency Statistics Centre) [7]. DATA 1 is the number of passengers per month in the 7 years period used to train the NN. The number of data for CN-PSOGA algorithm training are 72 node (January 2006 until December 2011), and 24 node (January 2012 until December 2013) as the target to be predicted. DATA 2 uses the number of precipitation from 125 stations of BMKG (Agency for Meteorology, Climatology, and Geophysics) in 2003 [8]. The parameters used in CN-PSOGA algorithm for both data and methods for comparison are shown in Table 1.

Table 1. Parameters are used for CN-PSOGA

Method	Parameter	Value Parameter
Neural Network	Hidden layer	12 node
	Error tolerance	0,001
	learning rate	0,08
Genetic Algorithm	Population	Sum hidden layer
	Cromosom	5
	Maxiter	100
	Probability crossover	0,8
	Probability mutation	0,05
PSO	Velocity $c_1 = c_2$	2
	Inertia weight	1

This paper uses population as the number of hidden layer in the NN algorithm and maximum of iteration (maxiter) used is 100. These parameters are used to compare the minimum error (MSE), computation time, and epoch of CN-PSOGA.

2.1 Neural Network

Neural network is one of the optimization method that represents of the artificial human brain which always tries to stimulate the learning process of that human brain. The neural network is implemented using the computer program which can solve the number of calculation processes during the learning process [9]. The network consists of three layers: input, hidden, and output. Figure 1 shows neural network architecture with the first layer is input, the second layer is called hidden layers, and this network has one unit in the third layer which is called the output layer [10],[11].

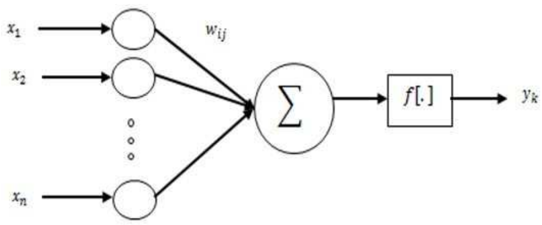

Figure 1 Architecture of Neural Network

$x_1, x_2, \ldots, x_n$ are input of the neuron with weights $w_1, w_2, \ldots, w_n$. f is an activation function, and y_k is output of the neuron signal. Tree layers neuron are connected by links which consists of weights.

2.2 Backpropagation

Backpropagation is the learning algorithm neural network. Backpropagation is a supervised learning method and this algorithm uses the technique of gradient search to minimize the error between the output and the real data [12]. This algorithm uses the output error to change the weights backwardly. The output of this network is then compared with the given target [13]. Backpropagation architecture is shown in Figure 2.

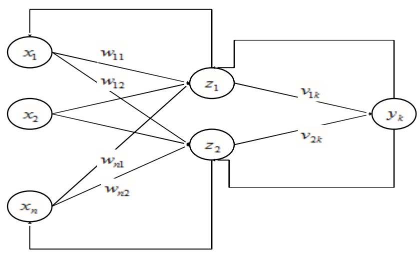

Figure 2 Architecture of Backpropagation

2.3 Conjugate Gradient

Conjugate gradient method is one of methods which is used to obtain the optimum value of the function. In general, the conjugate gradient method can be used to solve both the linear optimization and the nonlinear optimization problems [14].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Conjugate gradient method solves $\min_{x \in R^n} f(x)$ using iterative methods, i.e.

$$x_{t+1} = x_t + \alpha d_t$$

where x_t declares the results of iteration-t, α is steep size, and d_t represents the directions search. For example

$$g(x) = g(x_t),$$

then the search direction in first iteration uses the gradient descent direction i.e

$$d_0 = -g_0$$

afterwards the search direction is

$$d_{t+1} = -g_{t+1} + \beta_{t+1} d_t$$

which β_{t+1} is steep size in the conjugate gradient method with Fletcher-Reeves update [15].

2.4 Genetic Algorithm

Genetic algorithm is one of the search heuristic methods that mimics the process of evolution [4]. This algorithm can be used to carry out optimization problems. The process of obtaining the solution imitate of nature biological evolution mechanism such as replication, selection, crossover and mutation. Its implementation process includes encoding, generate population; fitness calculation; selection, replication; crossover; mutation and so on [4],[5].

The GA algorithm generates an initial population randomly. A new population will now develop from this initial population using three basic fundamental genetic processes. These are selection based on fitness, crossover and mutation. In each generation, the new population is evaluated using fitness function. In the selection process, the high fitness chromosomes are used to eliminate low fitness chromosomes. Some methods such as Roulette-wheel selection, Boltzmann selection, Tournament selection, Rank selection and Steady-state selection, are commonly used for reproduction or selection. However, selection alone does not produce any new individuals into the population [16]. Next comes to the crossover which is followed by mutation process. Crossover is the process that two member of the chromosome with high fitness values exchange genes. Applying this process to initial population will produce a new population. The final process is mutation where an individual should be selected randomly from the population and then changes the value of a gene with a low probability. The commonly used methods for mutation operations are simple mutation, uniform mutation, non-uniform mutation, Gaussian mutation and so on [5].

2.5 Particle Swarm Optimization

PSO method was first introduced by Kennedy and Eberhart in 1995. PSO is inspired from the biological behaviour of animal living in groups for example bee swarming, bird flocking and fish schooling [3],[4],[17],[18]. The basic idea of PSO algorithm is based on the movement and cooperation of birds in searching food. The algorithm consists of some particles where each individual particle keeps track of its position, velocity and best position. Each particle will move depending on the velocity and position with corresponding of two following equations:

$$v_{ik}^{t+1} = w.v_{ik}^t + c_1.r_1.(pb_{ik}^t - x_{ik}^t) + c_2.r_2.(gb_{g1}^t - x_{ik}^t)$$

$$x_{ik}^{t+1} = x_{ik}^t + v_{ik}^{t+1}$$

where v_{ik}^t and x_{ik}^t is the velocity and position of the j dimension of the i bird at the t time; pb_{ik}^t is the pbest and gb_{g1}^t is the gbest; w is inertial weight; c_1 and c_2 are coefficients of velocity; r_1 and r_2 are the random values [0,1] [3].

3. CN-PSOGA ALGORITHM

Every algorithm has advantages and disadvantages as shown from previous description. The conjugate gradient method can change every learning rate in neural network so that it can increase the rate of convergence. PSO is an algorithm based on the best particle in its population. Because of the particles of PSO are gathered to the best position, the best position particles can obtain the solution better in solving problem of optimization. The genetic operators, such as crossover and mutation, are used to find the solution of optimization. PSO with its calculation velocity can get solutions quickly by adding some GA operators (crossover and mutation) which act to expand the solution search area [19]. The PSO and GA algorithm are expected to give additional weight to neural network to obtain minimum error and speed up the rate of computation time. Flowchart of CN-PSOGA is shown in Figure 3.

The steps of CN-PSOGA to solve the training data can be explained as follows:

Step 1 : Input the training data.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Step 2 : Initialize all of weights used as the starting value of feedforward method.

Step 3 : Update *the* velocity rate in NN using CGFR method, and input the parameter of CGFR method $(\eta, d_t, g_t, \text{and}\, \beta_t)$.

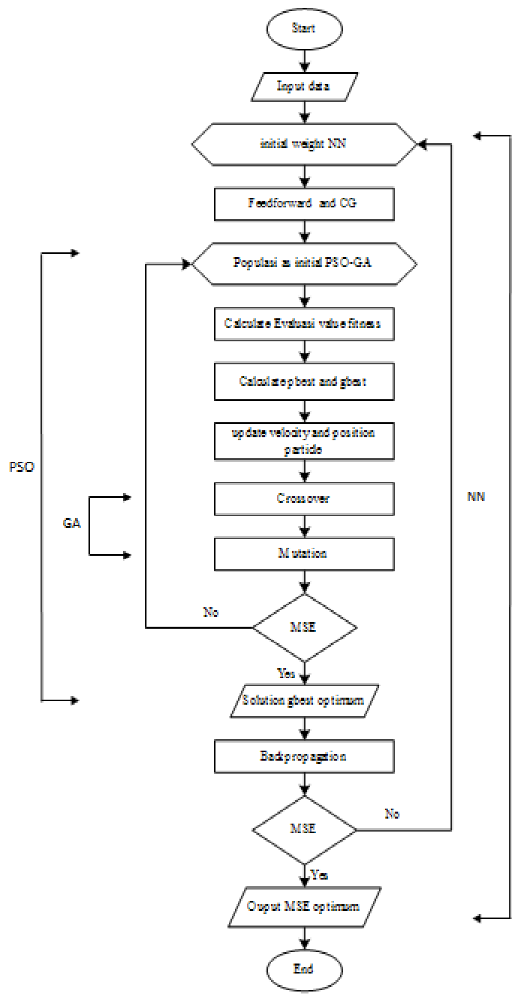

Figure 3 Flowchart CN-PSOGA

Step 4 : Optimize PSO and GA to obtain initial parameter: $x, v, w, c_1, c_2, r_1, r_2$, and maxiter.

Step 5 : Calculate the fitness value of each individual.

Step 6 : Determine the pbest and the gbest.

Step 7 : Update the velocity and the position of PSO particle.

Step 8 : Calculate the crossover process on the individual which is already generated and calculate the mutation process.

Step 9 : Repeat the 3-8 steps until the stopping condition is completed, i.e. until the value of iteration is equal to maxiter and generated the gbest solution.

Step 10 : The value of gbest is used to change the δ factor on the feedforward method

Step 11 : The weight generated from PSO and GA is used on backpropagation method

Step 12 : Repeat 3-11 steps until stopping condition is completed.

4. RESULT

In this section, we compare the proposed method with three other methods, namely, NN, CG-NN and NN-GA using DATA 1 and DATA 2. Using parameters as in Table 1, after 20 times of running program on algorithm using Matlab 2013b, Intel Core i3 2.1 GHz Duo, 4 GB of memory and OS Windows 7, get the results as in Table 2 and Table 3. These tables show that the addition of conjugate gradient method on the CN-PSOGA can decrease the minimum error and the required computation time than NN-GA which has a long computation time. NN and CG-NN has a short computation time, but these NN and CG-NN have a greater mean square error (MSE) than CN-PSOGA. These indicate that the CN-PSOGA has the best fitting to data when compared with other methods.

Based on the comparison Table 2 and Table 3, it is shown that the purpose of finding the error value is to see the closeness of the value generated by the proposed methods. The smaller error value of that method indicates that the fitting results are closer to the real data. The comparison of the number of iterations by each method to complete each trial data shows that NN and CG-NN have nearly the same number of iteration, computation time and MSE value, but CN-PSOGA method has the smallest MSE value compared with other methods, although this proposed method has a longer computational time. It can be concluded that the CN-PSOGA method on DATA 1 and DATA 2 is able to work well and the results obtained are almost equal to 20 times the running program.

In Figure 4, area which is marked with a circle is used to enlarge the area in order to see the comparison chart. This figure shows about comparison between the training results of CN-PSOGA algorithms and other algorithms using DATA 1. The chart of DATA 2 is similar to the DATA 1 (chart is not shown). The result shows that algorithm CN-PSOGA can generate the closest solutions to the real data. Figure 4 shows that CN-

Journal of Theoretical and Applied Information Technology
10th July 2015. Vol.77. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PSOGA methods is able to get the best solution of other methods, but based on Table 1 and Table 2, the computation time CN-PSOGA method is still longer than the CG-NN and NN.

Comparison of error value can be seen in Figure 5. The area which is marked with a circle is used to enlarge the area in order to see the comparison chart. The enlarge area is represented in Figure 5 on the right. The purpose of comparative value of error is to see the closeness results generated by the NN, CG-NN, NN-GA, and CN-PSOGA method. The smaller error value of the method, the closer results is to the real data. Figure 5 shows that CN-PSOGA methods is the best method in minimizing error of other methods, but this method requires the longer computation time than NN and CG-NN methods.

Table 2. Results of Running Program for Method of DATA 1

No	Method	MSE	Time (s)	Epoch
1	NN	0.0618	6.5236	8
2	CG-NN	0.0564	5.6488	8
3	NN-GA	5.7479×10^{-5}	1530.5	21
4	CN-PSOGA	8.5664×10^{-8}	21.1961	8

Table 3. Results of Running Program for Method of DATA 2

No	Method	MSE	Time (s)	Epoch
1	NN	0.060548	7.8905	6
2	CG-NN	0.058296	7.8315	6
3	NN-GA	6.4175×10^{-5}	1167.7	13
4	CN-PSOGA	1.0793×10^{-7}	23.148	6

5. CONCLUSION

In this paper, we have introduced the NN algorithm with CG, and the combination of PSO and GA as weight initialization. This method is compared with three other methods. Based on simulation of DATA 1 and DATA 2, CN-PSOGA method is able to obtain a better approximation results than the NN, CG-NN, and NN-GA methods. It can be concluded that the CN-PSOGA method has the best solution of other methods. Computation time is more efficient when compared with NN-GA method. This is because the solution of the proposed method with conjugate gradient does not require the initial value of velocity rate in NN. Furthermore, PSO-GA method provides the solution to add the weights for the neural network methods.

CN-PSOGA method has several advantages, such as is in accordance with the data with irregular fluctuations and has high performance for data very much.

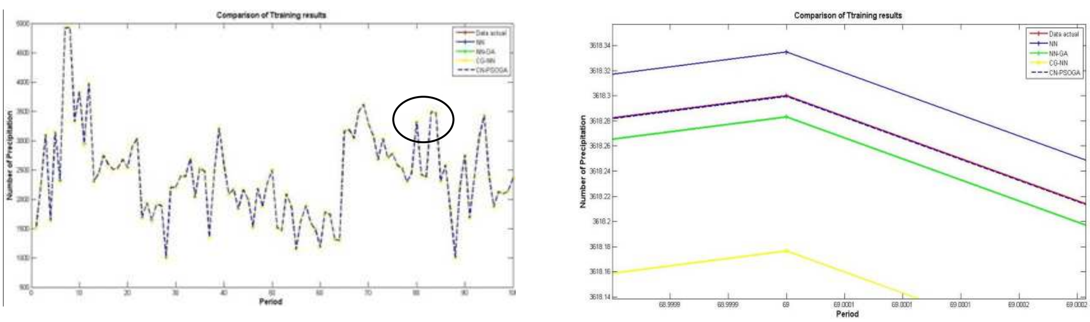

Figure 4 Comparison of Training results

Journal of Theoretical and Applied Information Technology
10th July 2015. Vol.77. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

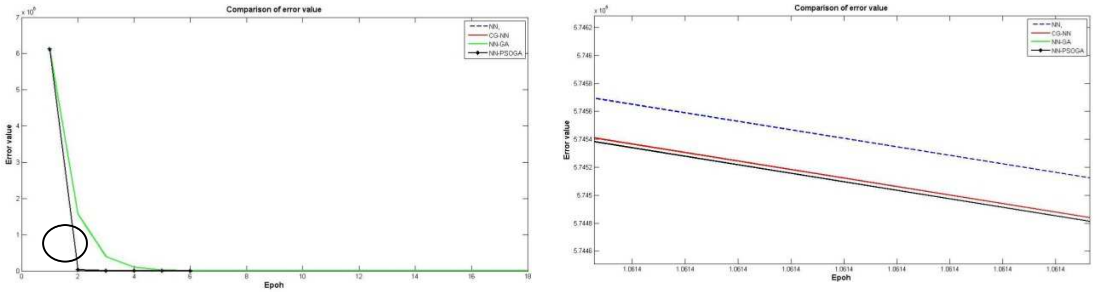

Figure 5 Comparison of error value

However, CN-PSOGA has several drawbacks including network architecture that is trial and error so it contains a lot of possibilities. The training process takes a long time compared to the standard method and has constraints in determining the number of nodes according to the data.

In future work, the development of the CN-PSOGA will include differentiation techniques to eliminate non-stationary patterns, techniques that can determine the number of nodes in the input, investigation of additive models to improve performance and overcome some of the weaknesses of modelling time series with missing data.

REFRENCES:

[1] N. M. Nawi, M. R. Ransing, and R. S. Ransing. “An Improved Learning Algorithm based on the Conjugate Gradient Method for Back Propagation Neural Networks”. *World Academy of Science, Engineering and Technology*. Vol. 20, No. 446, 2008, pp. 905-908.

[2] A. Y. Al-Bayati, N. A. Sulaiman, and G. W. Sadiq. ”A Modified Conjugate Gradient Formula for Backpropagation Neural Network Algorithm”. *Jurnal of Cumputer Science*. Vol.5, No.11, 2009, pp. 849-856.

[3] S. Labed, A. Gherboudj, and S. Chiki. “A Modified Hybrid Particle Swarm Optimization for Solving the Traveling Salesman Problem”. *Journal of Theoretical and Applied Information Technology*. Vol. 34, No. 02, 2011, pp. 11-16.

[4] R.K. Jatoth, A. K. Jain, and T. Phanindra. “Liquid Level Control of Three Tank System Using Hybrid GA-PSO Algorithm”. *Nirma University International Conference on Engineering* (NUiCONE), 2013, pp. 1-7.

[5] P. Li, Z. TAN, L. YAN, and K. DENG. “Time Series Prediction of Mining Subsidence Based on Genetic Algorithm Neural Network”. *International Synposium on Computer Science and Society*, 2011, pp. 83-86. DOI 10.1109/ISCCS.2011.30.

[6] D. Singh, V. Singh, and U. Ansari. “Binary Particle Swarm Optimization with Crossover Operation for Discrete Optimization”. *International Journal of Computer Applications*. Vol. 28, No.11, 2011, pp. 0975-8887.

[7] BPS,”Agency Statistics Centre”. http://www.bps.go.id/tab_sub/view.php?kat=2&tabel=1&daftar=1&id_subyek=17&notab=8 accessed October 13, 2014, at 8:44 pm, in Indonesian.

[8] BMKG, “Agency for Meteorology, Climatology, and Geophysic)”, http://www.bps.go.id/tab.sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=151&notab=3 accessed October 13, 2014, at 7:44 pm, in Indonesian.

[9] W. Huang, H. Hong, G. Song, and K. Xie. “Deep Process Neural Network for Temporal Deep Learning”. *International Joint Conference on Neural Networks* (China). July 6-11, 2014, pp. 1484-4799.

[10] K. G. Upadhyay, A. K. Choudhary, and M. M. Tripathi. “Short-term Wind Speed Forecasting using Feed-forward Backpro- pagation Neural Network”. *International Journal of Engineering, Science and Technology*, Vol. 3, No. 5, 2011, pp. 108-109.

[11] A.R. Naik and S. K. Pathan. ”Weather Classification and Forecasting Using Backpropagation Feed-forward Neural Network”. *International Journal of Scientific and Research Publication.* Vol. 2, No. 12, 2012, pp. 1-2.

[12] Ch. J. Devi, B. S. P. Reddy, K. V. Kumar, B. M. Reddy, and N. Raja. “ANN Approach for Weather Prediction using Backpropagation”. *International Journal of Engineering Trends and Technology.* Vol. 3, No. 1, 2012. pp. 19-22.

[13] D. A. Kumar, and S. Murugan. “Performance Analysis of MLPFF Neural Network Back propagatioan raining Algorithm for Time

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Series Data". *World Congress on Computing and Communication Technologies*., 2014, pp. 114-119. DOI 10.1109.

[14] M. Silcowitz-Hansen, S. Niebe, and K. Erleben. "A Nonsmooth Nonlinear Conjugate Gradient Method for Interactive Contact Force problem". *Vis Comput* 26, 2010, pp. 893-901. DOI 10.1007/s00371-010-0502-6

[15] O. Kisi and E. Uncuoglu. "Comparison of Three Backpropagation Training Algorithm for Two Case Studies". *Indian Journal of Engineering & Material Sciences.* Vol. 12, October 2005, pp. 434-442.

[16] A.G. Karegowda, A. S. Manjunath, and M. A Jayaram. "Application of Genetic Algorithm Optimized Neural Network Connection Weights for Medical Diagnosis of Pima Indians Diabetes". *International Journal on Soft Computing* (IJSC). Vol. 2, No. 2, 2011, pp. 16-17.

[17] F. Hembecker, H. S. Lopes, and W. Godoy Jr. "Particle Swarm Optimization for the Multidimensional Knapsack Problem". *B. Beliczynski et al. (Eds.):ICANNGA 2007, Part I, LNCS 4431*, pp. 358-365.

[18] J. Chang and X. Xu. "Applying Neural Network with Particle Swarm Optimization for Energy Requirement Prediction". *World Congress on Intelligent Control and Automation.* June 25-27, 2008, pp. 6161-6162.

[19] F. A. Tisna, S. Abusini, and A. Andari. "Hybrid Greedy-article Swarm Optimization-Genetic Algorithm and Its Convergence to Solve Multidimensional Knapsack Problem 0-1". *Journal of Theoretical and Applied Information Technology*. Vol. 58, No. 3, 2013, pp. 522-529.