Optimization of Performance of Genetic Algorithm for 0-1 Knapsack Problems Using Taguchi Method

A.S. Anagun and T. Sarac

Eskisehir Osmangazi University, Industrial Engineering Department,
Bademlik 26030, Eskisehir, Turkey
{sanagun, tsarac}@ogu.edu.tr

Abstract. In this paper, a genetic algorithm (GA) is developed for solving 0-1 knapsack problems (KPs) and performance of the GA is optimized using Taguchi method (TM). In addition to population size, crossover rate, and mutation rate, three types of crossover operators and three types of reproduction operators are taken into account for solving different 0-1 KPs, each has differently configured in terms of size of the problem and the correlation among weights and profits of items. Three sizes and three types of instances are generated for 0-1 KPs and optimal values of the genetic operators for different types of instances are investigated by using TM. We discussed not only how to determine the significantly effective parameters for GA developed for 0-1 KPs using TM, but also trace how the optimum values of the parameters vary regarding to the structure of the problem.

1 Introduction

The KP is a well-known combinatorial optimization problem. The classical KP seeks to select, from a finite set of items, the subset, which maximizes a linear function of the items chosen, subject to a single inequality constraint.

KPs have been mostly studied attracting both theorists and patricians. 0-1 KP is the most important KP and one of the most intensively studied discrete programming problems. The reason for such interest basically derives from three facts: (a) it can be viewed as the simplest integer linear programming problem; (b) it appears as a sub problem in many more complex problems; (c) it may represent a great many practical situations [1].

Many optimization problems are combinatorial in nature as 0-1 KP and quite hard to solve by conventional optimization techniques. Recently, GAs have received considerable attention regarding their potential as an optimization technique for combinatorial optimization problems [2-5].

In this paper, we present an approach for determining effective operators of GAs (called design factors) for solving KP and selecting the optimal values of the design factors. The KP is defined as follows:

$$\max_{x} \left\{ \sum_{j=1}^{n} p_j x_j \,\middle|\, \sum_{j=1}^{n} w_j x_j \le c,\, x \in [0,1]^n \right\} \tag{1}$$

M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 678 – 687, 2006.

where n items to pack in some knapsack of capacity c. Each item j has a profit p_j and weight w_j, and the objective is maximizing the profit sum of the included items without having the weight sum to exceed c.

Since the efficiency of GAs depends greatly on the design factors, such as; the population size, the probability of crossover and the probability of mutation [6], selecting the proper values of the factors set is crucial. Various reproduction and crossover types can be used in regard to optimizing the performance of the GA. In addition, operator types may affect the efficiency of GA. In this study, to investigate the affect of the factors briefly discussed, we generated nine instances composed of three different sizes (*50, 200* and *1000*) and three different types correlation (*uncorrelated, weakly correlated* and *strongly correlated*). To optimize the efficiency of GAs in solving a 0-1 KP, we examined closely which genetic operators must be selected for each instance by using TM.

The paper is organized in the following manner. After giving brief information about 0-1 KP, the GA for 0-1 KPs and the parameters that affect the performance of the GA are introduced. The fundamental information about TM is described next. Then the steps of experimental design are explained and applied to the problems taken into account. The results are organized regarding to the problems. Finally, the paper concludes with a summary of this study.

2 Genetic Algorithm for 0-1 Knapsack Problems

GAs are powerful and broadly applicable in stochastic search and optimization techniques based on principles from evolution theory [7]. GAs, which are different from normal optimization and search procedures: (a) work with a coding of the parameter set, not the parameters themselves. (b) search from population of points, not a single point. (c) use payoff (objective function) information, not derivatives or other auxiliary knowledge. (d) use probabilistic transition rules, not deterministic rules [8].

KPs that are combinatorial optimization problems belong to NP-hard type problems. An efficient search heuristic will be useful for tackling such a problem. In this study, we developed a GA (KP-GA) working with different crossover operators (1:*single-point*,2:*double-point*,3:*uniform*) and reproduction types (1:*roulette wheel*,2:*stochastic sampling*, 3:*tournament*) to solve the 0-1 KP. The KP-GA is coded with VBA. The proposed GA for solving 0-1 KPs is discussed below.

2.1 Coding

We shall use an n bit binary string which is a natural representation of solutions to the 0-1 KPs where one means the inclusion and zero the exclusion of one of the n items from the knapsack. For example, a solution for the 7-item problem can be represented as the following bit string: [0101001]. It means that items 2, 4, 7 are selected to be filled in the knapsack. This representation may yield an infeasible solution.

2.2 Fitness

The fitness value of the string is equal to the total of the selected items profit. To eliminate the infeasible string we use penalty method. If any string is not feasible, its fitness value is calculated by using the procedure penalty is given below:

```
procedure penalty:
  if ∑_{i=1}^{n} w_i >c then;
        begin
        p:= 1-( ∑_{i=1}^{n} w_i /5*c);
        if p <= 0 then p:= 0.00001;
        fitness value of string i= fitness value of string i*p;
        end;
```

2.3 Genetic Operators

A classical GA is composed of three operators: reproduction, crossover and mutation. Operators taken into consideration for solving 0-1 KP problems with GA are discussed as follows.

The reproduction operator allows individual strings to be copied for possible inclusion in the next generation. The chance that a string will be copied is based on the string's fitness value, calculated from a fitness function. We use three types of reproduction operators; roulette wheel, remainder stochastic sampling, and tournament.

Crossover enables the algorithm to extract the best genes from different individuals and recombine them into potentially superior children. KP-GA has three different crossover operators; single point, double point and uniform.

Reproduction and crossover alone can obviously generate a staggering amount of differing strings. However, depending on the initial population chosen, there may not be enough variety of strings to ensure the GA searches the entire problem space, or the GA may find itself converging on strings that are not quite close to the optimum it seeks due to a bad initial population. Some of these problems may be prevented by introducing a mutation operator into the GA. In KP-GA, a random number is generated for all genes. If the random number is smaller than mutation rate, the value of the gene is changed. If it is 0, its new value will be 1 and if it is 1, it will be 0, respectively. The value of gene is protected, if random number is bigger than the mutation rate.

When creating a new generation, there is always a risk of losing the most fit individuals. Using elitism, the most fit individuals are copied to the next generation. The other ones undergo the crossover and mutation. The elitism selection improves the efficiency of a GA considerably, as it prevents losing the best results.

The termination condition is a check whether the algorithm has run for a fixed number of generations. The number of generations of 1000 is selected as termination condition for all of the experiments conducted.

3 Taguchi Method

In principle, Taguchi design of experiments is used to get information such as main effects and interaction effects of design parameters from a minimum number of experiments. The behavior of a product or process is characterized in terms of design (*controllable*) and noise (*uncontrollable*) factors. Thus, TM may be applied to determine the best combination of design factors and to reduce the variation caused by the noise factors [9].

The TM uses a special design of orthogonal arrays (OAs) to study the entire parameter space with only a small number of experiments. Experimental design using OAs, recommended by Taguchi, not only minimizes the number of treatments for each trial, but also keeps the pair-wise balancing property [10].

An OA is basically a matrix of rows and columns in which columns are assigned to factors or their interactions and rows represent the levels of various factors for a particular experimental trial [11]. The treatment combinations are chosen to provide sufficient information to determine the factors' effects using the analysis of means. The OA imposes an order on the way the experiment is carried out. Orthogonal refers to the balance of the various combinations of factors so that one factor is given more or less weight in the experiment than the other factors. Orthogonal also refers to the fact that the effect of each factor can be mathematically assessed independently of the effects of the other factors [12].

The first step of designing an experiment with known numbers of factors in Taguchi's method is to select a most suitable OA, which design to cover all the possible experiment conditions and factor combination. The selection of which OA to use predominantly depends on these items in order of priority [13]: (1) the number of factors and interactions of interest, (2) the number of levels for the factors of interest, and (3) the desired experimental resolution or cost limitations.

In order to select an appropriate OA for experiments, the total degrees of freedom need to be computed. The degrees of freedom are defined as the number of comparisons between design factors that need to be made to determine which level is better and specifically how much better it is [14]. While the degrees of freedom associated with a design factor is one less than the number of levels, the degrees of freedom associated with interaction between two design factors is given by the product of the degrees of freedom for the two design factors. Basically, the degrees of freedom for the OA should be greater than or at least equal to those for the design factors.

Once the levels of design factors are settled, the analysis of means (ANOM) is conducted to find affection of each factor on the performance criterion by calculating the mean of entire data of the design factors. Hence, the optimum level of each design factor can be found by concentrating on its corresponding response graph. The analysis of variance (ANOVA) is then performed to determine the significant factors for the selected criterion. Finally, a prediction model consisting of the significant factors is built and confidence intervals for estimated mean and each of the significant factors are constructed.

4 The Experimental Design

In this paper, TM is applied to search the optimum values of the parameters of GAs designed for 0-1 KPs which are generated in terms of the number of items and correlation among weights and profits of items, respectively. With this consideration, it is aimed that whether the optimum values of the parameters affecting the performance of GAs may be changed while the numbers of items and correlation among weights and profits of items increasing and/or decreasing.

Five design factors are identified as potentially important for performance of GAs: (1) population size, (2) crossover rate, (3) mutation rate, (4) crossover type, and (5) reproduction type. For each of the design factors, based on the related research, three possible levels were considered as: population size of (10, 30, 50), crossover rate of (0.60, 0.75, 0.90), mutation rate of (0.001, 0.005, 0.01), crossover type (1, 2, 3) and reproduction type of (1, 2, 3), respectively.

L_{27} (3^{13}) OA was selected since it is the most suitable plan for the conditions being investigated, which allows for examining 13 three-level design factors and/or interactions among them with 27 trials. Regarding to the light of the preliminary tests, the selected factors and the interactions between design factors were assigned to the columns of the OA before conducting tests. For further information on OAs and assigning factors to an OA, readers may refer to [9].

In order to observe the effects of noise sources, each experiment was repeated three times (3*27=81) under the same conditions. The order to the experiments was made random in order to avoid noise sources that had not been taken into account initially and that could affect the results in a negative way. To investigate how KP-GAs behave for different 0-1 KPs, nine types of data instances were randomly generated regarding to the number of items and the correlation among weights and profits of items. The L_{27} OA was then applied to each of the KPs and the optimum levels of the design factors that affect the performance of the KP-GA were determined separately.

5 Results

Based on the combinations of design factors assigned to the selected OA, twenty-seven different KP-GAs are formed for each of the 0-1 KPs. Three types of instances representing the correlation among weights and profits of items are generated by using the following definitions and Eq. (2) that are proposed by [15]:

Uncorrelated instance: the weights w_j and the profits p_j are uniformly random distributed in [1, R], R=1000.
Weakly correlated instance: the weight w_j are distributed in [1, R] and the profits p_j in [w_j-R/10, w_j+R/10] such that p_j•1.
Strongly correlated instance: the weights w_j are distributed in [1, R] and the profits are set to p_j=w_j+R/10.

$$\text{Capacity is chosen as } c=\frac{1}{10}\sum_{j=1}^{n} w_j \tag{2}$$

The data obtained from the trials; for instance, when the number of items is 50 and there is no correlation among instances, coded as *[50_Unc]*, are analyzed as follows.

5.1 Analysis of Mean (ANOM)

As mentioned in [16], the Taguchi has created the S/N (signal to noise) ratio to quantify the present variation. The term "*signal*" represents the desirable value (mean) and the term "*noise*" represents the undesirable value (standard deviation). There are several S/N ratios depending on the types of characteristics; lower the better (LB), nominal the best (NB), and higher the better (HB). Since the nature of the objective function for KPs is maximization, higher the better (HB) criterion was selected as a performance statistics:

$$Z_{HB} = -10\log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^{2}}\right] \quad (3)$$

where n is the number of repetition of simulation under the same condition of design factors, y the characteristics, and subscript i indicates the simulation number of design factors in the OA table. As shown in Eq. (3), the greater the S/N ratio, the smaller is the variance of performance measure around the desired value.

The data for *[50_Unc]* and Eq. (3) are applied to obtain the S/N response graph which is depicted in Fig.1. From Fig.1, the optimum levels of the design factors of KP-GA designed for *[50_Unc]*, A3B2C3D3E1 can be found and its corresponding values are shown as; A3: population size of 50, B2: crossover rate of 0.75, C3: mutation rate of 0.01, D3: crossover type of 3, and E1: reproduction type of 1.

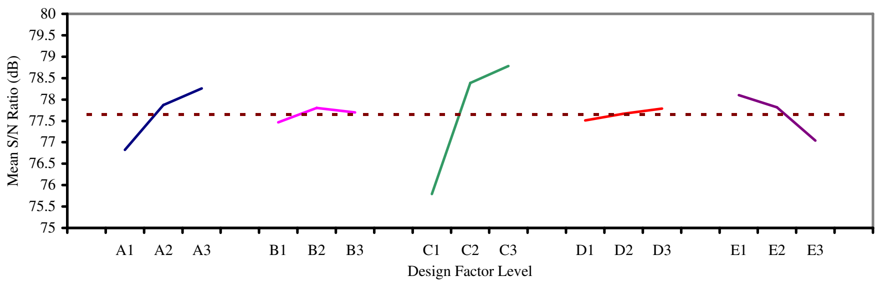

Fig. 1. S/N response graph for *[50_Unc]*

5.2 Analysis of Variance (ANOVA)

ANOVA, one of the most commonly and widely used analytical tools, is a technique that subdivides the total variation into meaningful components or sources of variation [17]. Using the ANOVA, one may be able to spot the key components (or factors) that cause excessive variation in products or processes. The ANOVA can be done with the raw data or with the S/N data.

The ANOVA based on the raw data signified the factors which affect the average response rather than reducing the variation. On the other hand, the ANOVA based on the S/N data takes into account both these aspects and so it was used here. Since all columns in OA are assigned with the factors and interactions, called saturated design, the variations due to error are estimated by pooling the estimates of the factors and interactions having least variance. This also helps in determining Fisher test (F-test) for finding out the confidence level of the results.

The ANOVA performed for the S/N data of *[50_Unc]* using ANOVA_TM software is given in Table 1. Table 1 indicates that factor C has the largest effect on the performance criterion. Factor A and factor E have the next largest effects, respectively. The remaining design factors have no effects due to their F values. Table 1 also indicates that the interaction of AxC has moderate effect and the interaction of CxE has least effect on the performance criterion. It is observed that the response graphs for the interactions' effects provided the identical results as the ones obtained regarding to the response graphs for the design factors. However, the interactions, in addition to the main design factors, should be considered to calculate the estimated mean S/N for the problem of *[50_Unc]*.

Table 1. ANOVA for S/N data of *[50_Unc]* – After pooling

Source	Pool	Df	S	V	F	S'	rho%
A	[N]	2	9.92062	4.96031	28.00694	9.56640	11.68
B	[N]	2	0.53491	0.26745	1.51008	0.18169	0.22
CxD	[Y]	4	0.56962	0.14240	-	-	-
CxE	[N]	4	2.26365	0.56591	3.19525	1.55521	1.90
C	[N]	2	47.46478	23.73239	133.99802	47.11056	57.50
AxC	[N]	4	14.52187	3.63047	20.49839	13.81343	16.86
BxC	[Y]	4	0.85650	0.21412	-	-	-
D	[Y]	2	0.34502	0.17251	-	-	-
E	[N]	2	5.45266	2.72633	15.39343	5.09844	6.22
(e)		10	1.77114	0.17711		4.60490	5.62
Total	[-]	26	81.92963	3.15114			

The column under the rho% gives an idea about the degrees of contribution of the factors to the performance criterion. To verify that the factors and levels selected for the experiment are reasonably correct, after pooling, the rho% accounted by the different factors cumulatively should be greater than 60% [18]. Since the total of rho% regarding to the significant design factors and interactions is approximately 94.4%, it may be said that the conducted experiment to find the optimal values of the design parameters for KP-GA applied to *[50_Unc]* problem are appropriate and furthermore, confirmation tests may be run.

5.3 Confirmation Tests

In order to predict and/or verify the improvement of the performance characteristic, confirmation tests should be conducted using the optimal levels of the design factors based on the results obtained from ANOM and ANOVA. The estimated S/N ratio η_{opt} using the optimal levels of the main design factors and interactions can be calculated as [19]:

$$\eta_{opt} = \hat{\eta} + \sum_{j=1}^{o} (\overline{\eta}_j - \hat{\eta}) \tag{4}$$

where $\hat{\eta}$ is the total mean S/N ratio, $\overline{\eta}_j$ the mean S/N ratio of the main design factor or interaction at the optimal level, and *o* the number of the design factors that affect the performance characteristic. In regard to optimal levels of the design factors that significantly affect the performance of KP-GA used for *[50_Unc]*, the estimated S/N ratio of 78.94190 (dB) is computed using Eq. (4).

The confidence interval of the estimated mean S/N ratio can be calculated by considering the following equation to verify whether the optimal solution of the objective function for the problem of *[50_Unc]* or targeted value is reached [20]:

$$CI = \sqrt{\frac{F(\alpha,1,v_e)\, V_e}{n_{eff}}}\ ;\ n_{eff} = \frac{N}{1+\sum(\cdot)} \tag{5}$$

where $F(\alpha,1,v_e)$ is the F-ratio required for $\alpha = \text{risk}$, v_e the degrees of freedom of error, V_e the pooled error variance, n_{eff} the effective sample size, N is the total number of trials, and $\sum(\cdot)$ the total degrees of freedom associated with items used in η_{opt} estimate. A confidence interval of 95% for the estimated S/N, $F(0.05,1,10) = 4.96$, $V_e = 0.17711$, and the effective sample size is $n_{eff} = 1.8$. Thus, the 95% confidence interval of the estimated optimum (dB) is computed as [78.24330; 79.64049].

The optimum levels of the design factors obtained for the problem of *[50_Unc]* are applied to the other 0-1 KPs generated with respect to the number of items and the correlation among weights and profits of items to examine whether the levels determined for the problem of *[50_Unc]* by means of TM are appropriate for the remaining generated 0-1 KPs. It is observed that the optimum levels of the design factors are basically problem dependent meaning that the appropriate levels of the design factors should be determined for each of the 0-1 KPs separately. Hence, each of the problems generated are analyzed based on the steps given in Section 5. In addition to results of problem *[50_Unc]*, the results obtained from analyses for the remaining 0-1 KPs being solved by KP-GA are shown in Table 2.

As shown in Table 2, the estimated S/N means for the generated 0-1 KPs are somehow different than the optimal solution obtained using LINGO, software for optimization modeling, although the constructed confidence levels included the optimal solutions for each of the problems concerned.

For a validity check, all of the experiments are repeated with the best combinations by considering different number of generations to verify whether the optimal solutions for the problems may be obtained. These experimentations concluded that the best combinations determined by the TM are able to produce the solutions as close as the ones obtained using LINGO, which makes the TM a powerful and sensitive approach for solving 0-1 KPs even if the size of the problems vary in terms of the numbers of items and the correlation among weights and profits of items.

Table 2. Results for analyses of generated 0-1 KPs

Problem Type: Number of items	Problem Type: Correlation	Treatment Combination	Estimated S/N Mean	Confidence Interval (95%)	Converted S/N*
50	No	A3B2C3D3E1	78.94190	[78.24330; 79.64049]	78.9271
50	Weak	A3B3C3D1E1	68.97514	[68.70190; 69.24840]	68.9307
50	Strong	A3B3C3D3E1	72.02041	[71.73480; 72.30610]	71.9209
200	No	A2B1C2D1E3	90.53733	[90.38260; 90.69206]	90.6665
200	Weak	A3B3C3D1E2	81.39253	[81.15990; 81.62520]	81.6133
200	Strong	A3B3C3D3E2	83.96448	[83.65620; 84.27280]	83.9770
1000	No	A3B1C1D1E2	104.12979	[103.2520; 105.0076]	105.0733
1000	Weak	A3B1C1D1E2	95.11788	[94.93680; 95.29900]	96.0850
1000	Strong	A3B1C2D1E2	97.53570	[97.36880; 97.70260]	98.2418

(*) optimal solution of the 0-1 KPs obtained from LINGO.

6 Conclusion

The aim of this study is not only to determine the optimum parameters that affect the performance of a GA designed for a combinatorial optimization problem, but also to investigate whether or not the optimum parameters of a GA applied to the problems, specifically 0-1 KPs formed in terms of the numbers of the items and the correlation among weights and profits of items, vary. The results obtained from the experiments conducted are summarized below.

The TM may be used to designate appropriate combinations of the parameters of the GA such that the optimal solutions may be reached; exactly the same solution for small and medium size problems, as close as the solution for the large size problems regardless of the correlation among weights and profits of items, comparing with the optimal solution obtained using LINGO.

The optimum value of 50 is obtained for the population size regardless of the structures of the problems concerned. The mutation rate is decreased as the size of the problem increased.

The double point crossover operator is not significant for the 0-1 KPs generated with respect to items and correlation among weights and profits of items. On the other hand, the appropriate crossover operator is obtained as uniform for small size problems, while single point crossover operator for large size problems despite the correlation.

The stochastic sampling reproduction operator seems appropriate for medium and large size problems, while roulette wheel is suitable for small size problems. Since the combinations of the parameters change as the correlation among weights and profits of items shift, as well as the number of items, the correlation should be taken into consideration for solving such problems.

The optimum parameters of GA designed for solving 0-1 KPs are dependent on the structure of problem. Hence, the initial values of the parameters may be selected using the results of this study as a table look-up. In order to provide a general outline for the persons who are interested in solving such problems by means of GA, the GA should be run for the different problems regarding to the numbers of items than the ones used in this study.

References

1. Martello, S., Toth, P.: Knapsack Problems Algorithms and Computer Implementations. John Wiley&Sons, England (1990)
2. Sakawa, M., Kato, K.: Genetic Algorithms with Double Strings for 0-1 Programming Problems. European Journal of Operational Research. 144 (2003) 581-597
3. Bortfeldt, A., Gehring, H.: A Hybrid Genetic Algorithm for the Container Loading Problem. European Journal of Operational Research. 131 (2001) 143-161
4. Taguchi, T., Yokota, T.: Optimal Design Problem of System Reliability with Interval Coefficient Using Improved Genetic Algorithms. Computers and Industrial Engineering. 37 (1999) 145-149
5. Michelcic, S., Slivnik, T., Vilfan, B.: The Optimal Cut of Sheet Metal Belts into Pieces of Given Dimensions. Engineering Structures. 19 (1997) 1043-1049
6. Iyer, S.K., Saxena, B.: Improved Genetic Algorithm for the Permutation Flowshop Scheduling Problem. Computers and Operations Research. 31 (2004) 593-606

7. Gen, M., Cheng, R.: Genetic Algorithms and Engineering Design. John Wiley&Sons, New York (1997)
8. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
9. Taguchi, G.: Systems of Experimental Design. Unipub Kraus International Publishers, New York (1987)
10. Georgilakis, P., Hatziargyriou, N., Paparigas, D., Elefsiniotis, S.: Effective Use of Magnetic Materials in Transformer Manufacturing. Journal of Materials Processing Technology. 108 (2001) 209-212
11. Antony, J., Roy, R.K.: Improving the Process Quality Using Statistical Design of Experiments: A Case Study. Quality Assurance. 6 (1999) 87-95
12. Fowlkes, W.Y., Creveling, C.M.: Engineering Methods for Robust Product Design. Addison-Wesley, Reading (1995)
13. Ross, P.J.: Taguchi Techniques for Quality Engineering. 2nd edn. McGraw-Hill, New York (1996)
14. Nian, C.Y., Yang, W.H., Tarng, Y.S.: Optimization of Turning Operations with Multiple Performance Characteristics. Journal of Materials Processing Technology. 95 (1999) 90-96
15. Martello, S., Pisinger, D., Toth, P.: New Trends in Exact Algorithms for the 0-1 Knapsack Problem. European Journal of Operational Research. 123 (2000) 325-332
16. Wu, D.R., Tsai, Y.J., Yen, Y.T.: Robust Design of Quartz Crystal Microbalance Using Finite Element and Taguchi Method. Sensors and Actuators. B92 (2003) 337-344
17. Roy, R.K.: A Primer on the Taguchi Method. VNR Publishers, New York (1999)
18. Ozalp, A., Anagun, A.S.: Analyzing Performance of Artificial Neural Networks by Taguchi Method: Forecasting Stock Market Prices. Journal of Statistical Research. 2 (2003) 29-45
19. Lin, T.R.: Experimental Design and Performance Analysis of TiN-coated Carbide Tool in Face Milling Stainless Steel. Journal of Materials Processing Technology. 127 (2002) 1-7
20. Syrcos, G.P.: Die Casting Process Optimization Using Taguchi Methods. Journal of Materials Processing Technology. 135 (2003) 68-74