International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA

Performance Evaluation of Selection Methods of Genetic Algorithm and Network Security Concerns

Hari Mohan Pandey[a]*

[a] Computer Science & Engineering, Amity University Uttar Pradesh, Noida, 201313, U.P., India

Abstract

Security is the prominent concern for the network and maintaining security is highly recommended. There exists several approaches has been attempted to address this challenging task. This paper presents the applicability of the genetic algorithm (GA) for security concerns. The working of the GA heavily depends on the various factors includes: reproduction operators, selection techniques, chromosome representation and problem type. There exist several selection methods presents play a vital role, but identifying the suitable one is a grand and persistent challenge. In this paper, a comparison of various selection techniques in the GA has been reported. The GA utilizes operators: crossover, mutation and selection to guide the searching in an iterative manner. A significant work has been conducted explains the importance of crossover and mutation probabilities, but very few researchers (some of them has shown the comparison of selection methods) presented the importance of selection approaches. The comparison of three: Rank based, Roulette wheel and Tournament selection techniques have been presented in this paper over Travelling Salesman Problem. Computational experiments have been conducted and results are collected considering distance should be as minimum as possible. Statistical tests (Paired T-Test and two ways ANOVA) are conducted to report the performance significance of selection techniques considered.

Keywords: Genetic algorithm; Travelling salesman problem; Selection methods.

* Corresponding author. Tel.: +91-9810625304
  *E-mail address:*profharimohanpandey@gmail.com

1. Introduction

Genetic Algorithms (GAs) are "adaptive heuristic search algorithms" works on the Darwin's principle of "survival of the fittest"[1]. GA has been applied in different fields of engineering includes machine learning, image processing, grammatical inference[16 17 18 19], natural language processing, language interpretation and others. Individual competes each other in a generation – one that succeeds passed to the next generation. The GA is a two-step process, namely selection of parents and applying reproduction operations on the selected parents[3]. Selection determines which individuals are chosen for the reproduction. It works on the approach: "the batter is an individual; the higher is its chance of being a parent[2]." There exist a number of approaches to select the parent chromosome that depends on the problem considered and its difficulty level. Identifying the most suitable selection method is considered as a crucial step in GA. The selection technique should be chosen cautiously, so that better individuals with high fitness value have a greater probability of selection. But the worst individuals should have little probability of selection and should not be completely discarded. This ensures that that our solution is global and reduces the risk of premature convergence[15].

In this paper, the comparison of different selection strategies for solving Travelling Salesman Problem (TSP) is presented. TSP is NP-Complete "combinatorial optimization problem[4]" in which our aim is to find the shortest route of a salesman i.e. travelling starting from his home city, covering each city in his list at least once and returning back to his home city[11]. There are various approaches to solve TSP like neural networks[8], simulated annealing[9], branch and bound[10], Genetic Algorithm[3 6] and many more. Since GAs is optimization search algorithms used for minimizing or maximizing certain function – is considered to solve TSP in our experiment. Various researchers have solved TSP using GA in[4 5 6 7]. Choosing a right selection technique is a very critical step in GA, since if not chosen correctly, it may lead to convergence of the solution to a local optimum.

The remaining paper is structured as: Section 2 presents the security concerns and how GA has been used. Section 3 gives an overview of GA for TSP. The selection strategies that we have analyzed are also explained in brief in the same section. Section 4 describes the simulation model in which experimental results are discussed and their analysis is done. The conclusion is drawn in section 5.

2. Genetic Algorithm and Security Concerns

Maintaining security is highly recommended to ensure safe and trusted communication, but still it is a challenging task to accomplish. It has been observed that communication over the Internet or over other network system suffers due to intrusions and misuse of the data. It motivated researchers to attempt this challenging task to achieve computer network security. There exist several approaches being employed for the intrusion detection, but unfortunately none of them are flawless. In the recent years, some encouraging results have been received, incorporating the GA. This section simply discusses the approaches developed for the network security where GA has been utilized.

Hoque et al.[20] has incorporated the GA to the information evolution to filter the traffic data leads to decrease in the complexity using KDD99 benchmark data sets. Banković et al.[21] has proposed a misused detection system, was based on the GA. The authors[21] have deployed a principle component analysis (PCA), which extract important features of the data. 22. Dutt and Chaudhuri[22] have used encryption/decryption heuristic GA to achieve the network security. Islam et al.[23] has discussed the importance of the GA for the network security and proposed a modified version of the GA uses special fitness function to detect security. This discussion leads to a conclusion that the power of the GA can be utilized for the network security and optimization purpose. The selection (survival selection, and parent selection) is a basic armory of the GA, greatly contributes to the success. Also, it is worth to mention that there are several selection techniques existing and most of the time difficult to pick the right one to perform the computational experiment, motivated the authors to conduct this study.

3. TSP uses Genetic Algorithm and Discussion of Selection Methods

This section presents an overview about the components of GA and their operations for solving TSP. GA uses a stochastic approach for randomly searching and optimizing the solutions. It ensures randomness and efficiency in

the search[3]. In GA, a chromosome shows a possible solution. The chromosome in TSP can be represented by the 'path' representation[11]. In TSP our main aim is to minimize the distance that we need to travel. Therefore, 'path' is the solution that we need to optimize and for that reason, it is represented as a chromosome in the GA process.

Algorithm-1: Procedure GA-TSP (No. Of cities)

Begin
Initialize GA and TSP parameters:
No. Of cities
Cities' coordinates
G_{max} shows the maximum number of generations
Size of the population
Crossover rate
Mutation rate
Tournament size
Generate random, initial population P (G)
Fitness ⟵ Evaluate P (G)
While (((Result is not Optimum) OR (Generation< G_{max}))
    Do select a couple of parent population P1 from P (G)
        Apply crossover to P1
        Apply mutation to P1
        G=G+1
        Update Population (P(G),P1(G))
End while
Display optimum result
End

The procedure for TSP using GA is explained using Algorithm-1 – starts by initializing GA's parameters like total generations, the size of the population, the size of the tournament and crossover, mutation probabilities. It also supplies important information like Number of cities and their coordinates. Then random population is generated and the fitness of each chromosome is calculated. A new generation is formed with the help of selection, crossover and mutation operators. The selection operator chooses two parents from the current generation, which then reproduce a new child with the help of crossover and mutation operators. This new child chromosome forms the next generation, which is better than the previous one. This process continues until an optimum solution is achieved, or generation reaches its maximum limit. A solution is said to be optimum if a certain percentage of the population (say 90%) have same optimum chromosome, out of which the best one is chosen as the optimum solution.

Three selection schemes: Roulette wheel, Rank based and Tournament selection are selected. Roulette Wheel Selection was proposed by Holland[12] assumed that the selection probability of an individual chromosome is directly related to the fitness. It works in a similar fashion as a roulette wheel, in which, the selection probability depends on the central angle of the roulette wheel. In the same way, in GA, a whole population is partitioned in different sectors and the selection probability of an individual (one sector) is represented as an individual's fitness to the total fitness of the population. The probability of selection of an individual I_i can be calculated using equation (1).

$$PS(I_i) = \frac{f(I_i)}{\sum_{i=1}^{n} f(I_j)} ; j = 1, 2, ..., n \tag{1}$$

Where 'n' is the population size and $f(I_i)$ is the fitness value of an individual I_i.

Linear Ranking Selection was proposed by Baker[13]. In this, an individual in a population is first sorted as per their fitness, and then assignment of the rank takes place. 'N' is the rank given to the best individual, whereas rank '1' is assigned to the worst. The probability of selection of an individual I_i is given in equation (2).

$$P_i = \frac{1}{N}(n^- + (n^+ - n^-)\frac{i-1}{N-1}); i \in \{1,...,N\} \tag{2}$$

Where P_i, $\frac{n^-}{N}$ and $\frac{n^+}{N}$ respectively denotes the selection probability of ith individual, worst individual and best individual.

Tournament Selection, the most popular selection techniques due to its less time complexity. In this, 'n' random individuals are chosen from the entire population and the individual with best fitness value is selected for the further processing of GA[14]. Number of individuals taking part in each tournament is known as tournament size.

4. Experiment Design, Results and Discussion

This section focuses on the experimental design, results that we collected and their comparative analysis. The implementations have been done using Net beans IDE 6.9 beta, with system configuration 64 bits Windows 8 Operating System, 4 GB RAM and Intel Core I5 1.80 GHz Processor. The performance for ten TSP samples (one sample corresponds to 10 test runs) is tested for 20, 40 and 60-cities TSP (test runs for each sample is shown in Table 1). For our experiments, we have taken a combination of ordering crossover and swap mutation. The crossover and mutation probability that have taken is 0.3 and 0.01 respectively. Tournament size used in Tournament selection is 5. Travelling distance has been considered as a result. All the experiments follow the same termination process, i.e. termination will not take place until the number of generations will reach to the threshold point. The threshold point is the maximum number of generations. From Table 1, we are unable to conclude that which selection technique provides the best solution to the considered problem. For this reason, paired sample T-Test was performed on our results to statistically analyze them and find the selection technique that is best suitable for our problem. The data were analyzed using computer-based statistical software package IBM SPSS statistics version 22. To start with analysis, first we need to define hypothesis outlined as:

Hypothesis:

H_0: $\mu_0 = \mu_1 = \mu_2$ (i.e. All the selection techniques give the same results)

H_A: At least one selection technique is different from the above.

Significance was set at 0.05 (95% confidence).

Table 2 shows the mean, Standard Deviation and Standard Mean Error of each selection technique. The selection techniques are paired together in three combinations, i.e. Pair 1 (Tournament and Roulette wheel selection), Pair 2 (Tournament and Rank based selection) and Pair 3 (Roulette wheel and Rank based selection).

Table 3 is the result of paired sample T-Test. In pairs sample T-test, the data is needed to be entered in pairs. In paired T-test, we are interested in knowing the difference between each observation. We apply the following approach: for each pair, calculate the difference and then conduct a one sample T-test on the difference. The 't' value can be calculated by applying equation (3).

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}} \tag{3}$$

Where $\overline{x}$ is mean, n is number of samples that are taken into consideration, i.e. 30, μ is the standard error mean and s is the standard deviation.

Table 1.
Outputs for different selection techniques for varying size of cities.

S.No.	Roulette wheel Selection			Tournament Selection			Ranking Selection		
	C=20	C=40	C=60	C=20	C=40	C=60	C=20	C=40	C=6

									0
1	120	258	462	146	243	474	135	221	447
2	123	247	461	124	233	470	114	233	440
3	124	278	454	152	259	494	134	240	442
4	115	273	478	155	274	483	123	262	382
5	161	276	483	125	279	451	125	238	434
6	123	275	443	147	286	513	125	220	422
7	135	304	457	147	284	461	136	257	429
8	117	265	432	126	241	468	116	265	439
9	122	281	489	153	271	432	115	252	441
10	140	232	393	119	299	464	114	242	438

C: City

Table 2.
Paired sample T-test

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Tournament	471.4700	10	8.29940	2.62450
	Roulette wheel	452.0200	10	9.66020	3.05482
Pair 2	Tournament	471.4700	10	8.29940	2.62450
	Ranking	446.5900	10	11.02970	3.48790
Pair 3	Roulette wheel	452.0200	10	9.66020	3.05482
	Ranking	446.5900	10	11.02970	3.48790

Table 3
Paired sample T-test (Sig. (2-tailed))

		Paired Differences					T	df	Sig.(2-tailed)
		Mean	Std. Deviation	Std. error	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Tournament - Roulette wheel	19.45000	13.24699	4.18907	9.97367	28.92633	4.643	9	.001
Pair 2	Tournament - Ranking	24.88000	13.52321	4.27642	15.20608	34.55392	5.818	9	.000
Pair 3	Roulette wheel - Ranking	5.43000	13.12572	4.15072	-3.95958	14.81958	1.308	9	.223

As we can see in Table 3, for Pair 1 (Tournament – Roulette Wheel), t (4.643) is not in its range (lower bound = 9.97367 and upper bound = 28.92633). This means that the null hypothesis H_0 is false for the Tournament and Roulette Wheel Selection. In Pair 2 (Tournament – Ranking), again t (5.818) is not in its range (lower bound = 15.20608 and upper bound = 34.55392). This means that the null hypothesis H_0 is false for Tournament and Ranking Selection also, whereas in Pair 3(Roulette Wheel - Ranking), t (1.308) lies within its range (lower bound = -3.95958 and upper bound = 14.81958). This means that the null hypothesis H_0 is true for Roulette Wheel and Ranking Selection. Therefore, we conclude that Rank based and Roulette Wheel Selection shows similar in their results, while other pairs are different from each other. Table 2 shows that Rank based selection have the minimum distance as compared to other techniques. Therefore, Rank based selection outperformed Roulette wheel and Tournament selection.

4. Conclusions

In our paper, we have done a comprehensive study of different selection techniques in GA to solve TSP. We then compared their performance in terms of the minimum distance required to get the shortest path in TSP. For our analysis purpose, we have taken 20, 40 and 60-city TSP with 10 samples for each selection scheme; each sample is

an average of 10 test runs. GA does not give the exact results, but gives most appropriate result for the desired problem. According to the experiments that we conducted, Rank based selection technique gave the best result in terms of distance. It was then followed by Roulette wheel and Tournament selection. We also concluded that Rank based and Roulette wheel selection shows similar in their results and can be paired together. GA can be applied to various NP-Complete problems like knapsack problem, 3-SAT problem, a subset problem, vertex cover problem, etc. In future, we can work on these NP complete problems and find the selection technique that is suited for each one of these problems. In this paper, we have specified the crossover and mutation methods and the reproduction parameters in order to get best results. In the future, we can do the same experiment for different combination of types of crossover and mutation techniques and/or change the values of reproduction parameters (p_c and p_m).

References

1. A. Shukla, H. M. Pandey, D. Mehrotra, "Comparative Review of Selection Techniques in Genetic Algorithm," Futuristic Trends in Computational analysis and Knowledge management (INBUSH ERA-2015), IEEE, Greater Noida, India, 2015, ISBN: 978-1-4799-8432-9.
2. Blickle, Tobias, and Lothar Thiele. "A comparison of selection schemes used in genetic algorithms." (1995).
3. Noraini, Mohd Razali, and John Geraghty. "Genetic algorithm performance with different selection strategies in solving TSP." (2011).
4. Tsujimura, Yasuhiro, and Mitsuo Gen. "Entropy-based genetic algorithm for solving TSP." Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES'98. 1998 Second International Conference on. Vol. 2. IEEE, 1998.
5. Sengoku, Hiroaki, and Ikuo Yoshihara. "A fast TSP solver using GA on JAVA."Third International Symposium on Artificial Life, and Robotics (AROB III'98). 1998.
6. Moon, Chiung, et al. "An efficient genetic algorithm for the traveling salesman problem with precedence constraints." European Journal of Operational Research 140.3 (2002): 606-617.
7. Carter, Arthur E., and Cliff T. Ragsdale. "A new approach to solving the multiple traveling salesperson problem using genetic algorithms." European journal of operational research 175.1 (2006): 246-257.
8. S. Bhide, N. John, and M. R. Kabuka. "A Boolean neural network approach for the traveling salesman problem," IEEE Transactions on Computers, 1993, Pp. 1271 – 1278. DOI: 10.1109/12.257714
9. Kirkpatrick, Scott, and Gérard Toulouse. "Configuration space analysis of travelling salesman problems." Journal de Physique 46.8 (1985): 1277-1292.
10. Finke, Gerd, Armin Claus, and Eldon Gunn. "A two-commodity network flow approach to the traveling salesman problem." Congresses Numeration 41.1 (1984): 167-178.
11. Larrañaga, Pedro, et al. "Genetic algorithms for the travelling salesman problem: A review of representations and operators." Artificial Intelligence Review 13.2 (1999): 129-170.
12. J.H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, MIT Press, 1992.
13. Baker, James Edward. "Adaptive selection methods for genetic algorithms. "Proceedings of an International Conference on Genetic Algorithms and their applications. 1985.
14. Goldberg, David E., and Kalyanmoy Deb. "A comparative analysis of selection schemes used in genetic algorithms." Urbana 51 (1991): 61801-2996.
15. H.M. Pandey, A. Choudhary, D. Mehrotra, A comparative review of approaches to prevent premature convergence in GA, in: Applied Soft Computing, 2014.
16. H.M. Pandey, Context free grammar induction library using genetic Algorithms, in: 2010 International Conference on Computer and Communication Technology (ICCCT), IEEE, 2010.
17. Choubey et al.," Developing genetic algorithm library using Java for CFG induction", Int. J. Adv. Technol. 2 (1) (2011) 117–128.
18. H.M. Pandey, A. Dixit, D. Mehrotra, Genetic algorithms: concepts, issues and a case study of grammar induction, in: Proceedings of the CUBE International Information Technology Conference, ACM, 2012.
19. H.M. Pandey, A. Choudhary, D. Mehrotra. "Grammar induction using bit masking oriented genetic algorithm and comparative analysis." Applied Soft Computing 38 (2016): 453-468.
20. Hoque, Mohammad Sazzadul, et al. "An implementation of intrusion detection system using genetic algorithm." arXiv preprint arXiv:1204.1336 (2012).
21. Banković, Zorana, et al. "Improving network security using genetic algorithm approach." Computers & Electrical Engineering 33.5 (2007): 438-451.
22. Dutt, I., S. Paul, and S. N. Chaudhuri. "Implementation of network security using genetic algorithm." Int J Adv Res Comput Sci Software Eng 3.2 (2013): 234-41.
23. Islam, A. B. M., et al. "Security Attack Detection using Genetic Algorithm (GA) in Policy Based Network." Information and Communication Technology, 2007. ICICT'07. International Conference on. IEEE, 2007.