Exploiting the Local Optima in Genetic Algorithm using Tabu Search

M. Sivaram[1], K. Batri[2], Amin Salih Mohammed[1] and V. Porkodi[1]

[1]Department of Information Technology, Lebanese French University, Erbil, Kurdistan Region, Iraq;
sivaram.murugan@lfu.edu.krd, kakshar@lfu.edu.krd, porkodisivaram@gmail.com
[2]Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology,
Kothandaraman Nagar, Dindigul – 624622, Tamil Nadu, India; Krishnan.batri@gmail.com

Abstract

Objectives: To explores the process of selecting retrieval schemes along with their weights, and fusion function for data fusion in information retrieval. **Methods/Statistical Analysis**: This has been carried out using the hybrid Genetic Algorithm. The fusion function, retrieval schemes and their weights lead to a tremendous combination. Finding an optimal solution from this great combination is entirely based on the exploration. **Findings**: We used, odd and even point crossover as an exploration tool. This exploration tool suffers a setback of slow convergence. The convergence rate can be improved by merging Tabu search, a best local search, with the genetic algorithm. This Tabu GA is used to select the retrieval schemes, weights and fusion function. The outcome of the experiments conducted over the test data sets namely: 1. adi, 2. cisi, and 3. cranlooks promising. We achieved 6.89% of improvement in performance, and the significance of the result is tested statistically. The convergence rate is also improved. **Application/Improvements:** We achieved 6.89% of improvement in performance, and the significance of the result is tested statistically. The convergence rate is also improved.

Keywords: Genetic Algorithm, Information Retrieval, Odd and Even Point Crossover, Tabu GA, Tabu Search

1. Introduction

Information Retrieval (IR) is the process of finding relevant information from the massive volume of data[1-3]. The IR system process, arrange, store, and proffer the relevant items based on the users' query. The correlation between the document and the query is calculated using various similarity measures[4]. The performance of the IR system is varying from one corpus to other[5-6]. Fusion is used to overcome this drawbacks[7-8].

Data fusion is a process merging results from more than one resources[9]. It combines the results from various retrieval schemes and strategies. The fusion function converts the multidimensional vector into a scalar[10].

The conversion of the vector to a scalar is varying from one function to another function. Hence, the problem becomes multi-objective. The optimization tools are useful for solving the multi objective issues[11]. GA is the best among these global optimization tools[12]. It has the bio-inspired operators such as reproduction, crossover and mutations for finding the optimal solution.

In our problem, we use GA for finding the best fusion function, retrieval schemes, and their weights. The problem demands more exploration, and we proposed methodology a new exploration tool called as *'odd and even point crossover'*. The exploration suffers a drawback of slow convergence, and it can be overcome by combining the best local search[13] algorithm with the GA. Tabu[14]

is one of the best local optimization tools. The Tabu GA is used to find the optimal solution for our problem.

The rest of the study is organized as follows. Section 2 gives the GA based selection and its results. Section 3 analyzes the convergence of GA based selection. The Tabu GA based selection is given in Section 4, and its convergence analysis, comparison with the GA based selection is given in section 5. Section 6 concludes with the future direction of our research.

2. Proposed Method

Fusion combines the advantages of all participating members[15,16]. In IR, fusion becomes a multi-objective optimization. It has to select the best fusion function, best retrieval schemes and the optimal weights for the retrieval schemes.

We used GA for fulfilling our need. As the optimization involves more than one goal and parameters, we have to explore the search space. '*Odd and Even Point crossover*' is used for this purpose, and it has been recorded in our previous work[17]. The experiments are conducted over three data sets namely adi, cisi, and cran. The characteristics of the data sets are given in Table 1.

The fusion function used in our experiments is COMB MNZ, COMB MAX, COMB MIN and COMB SUM[9]. The retrieval schemes used are: 1. dice, and 2. inner product and conjunctive normal form with P value as 1.5 and 2.5. The GA used here is to find the best fusion function, best retrieval schemes and the optimal weight for the selected schemes. Fitness Function used in our experiment is given below. Average 11 pt interpolated precision is used as the fitness function it is given by equation 1.

$$\text{Fit}(f) = \frac{\sum\limits_{i=1}^{n} \text{ave } P_s(q_i)}{N}$$

Where

N = Number of queries

P_s = Average precision of the query

We used 12 point string structure, and it is given by:

Table 1. Characteristics of datasets

Experiments	adi	cisi	cran
Number of documents	82	1460	1400
Number of terms	374	5743	4486
Number of queries	35	35	225
Average number of a document relevant to aquery	5	50	8
Average number of terms per document	16	45	56
Average number of terms per query	5	8	9

No of bits = 12: b0 ……..b11.

1. bo, b1, b2, Weight of the first retrieval strategy,

2. b3, b4, b5, Weight of the second retrieval strategy

3. b6, b7 - first retrieval strategy,

4. b8, b9 - second retrieval strategy,

5. b10, b11- Fusion function.

The gene position is divided into the odd and even location. We have a P_{odd} value. It is ranging from 0 to 1. If the randomly generated value is less than 0.5, then the crossover is carried out on odd location. If it greater than 0.5, then cross over carried out on even location. The experiments are carried out 50 times with the following GA parameter given in Table 2.The results obtained for the above parameter setting is given in Table 3. There is an improvement of 6.28% in the performance. It seems that the convergence is low and the variation is high. We want to test the convergence, and it is given in the next section.

3. Selection of Retrieval Scheme Using Ga

The odd and even point crossover is used as the robust exploration tool in comparison with the other existing crossover operators[18]. It successfully explored the search space. The algorithm and the results obtained using this operator is discussed in the previous section.

The previous section proffers the IR system performance and the tool properties are not discussed. This section is used to analyse the convergence property of the exploration tool namely odd and even point crossover. The selection problem for the data fusion in information retrieval has too many combinations. All the combinations ought to be explored. This need leads to the frequent disturbance in the gene structure and results in slow con-

Table 2.   GA parameters

Number of generations	G=100
Population Size	N=10
Cross over probability	$P_c=0.6$
Mutation Probability	$P_m=0.01$
Odd & Even Probability	$P_{odd}=0.5$

Table 3.   GA based selection results

Collection	Precision	% Improvement
ADI	0.3793	6.28
CISI	0.1872	3.42
CRAN	0.1418	3.35

vergence. The convergence property is analysed by using the following two considerations.

1. Average fitness value of the individual, and

2. Variation in the average fitness values from one to other generation.

3.1 Average Fitness Value of the Individual

The average value of the individuals present in each generation is recorded to measure the convergence property. This is continued for 100 generation as we set the number of generations as 100.

3.2 Variation in the Average Fitness Values from One Generation to Other

We want to confirm the slow convergence. It can be confirmed by measuring the difference in fitness value among two successive generations.

4. Tabu Ga

The genetic algorithm based selection has been discussed in section 3. The gene structure and the GA parameters are discussed. This section is used to discuss the Tabu GA for the data fusion. The Tabu list for the various components is given in Table 4. The Tabu tenure period is the restriction imposed over the individual components. During this period, the components are neither deleted nor included in the solution. The Tabu list and tenure periods are used as the exploitation tool.

4.1 Tabu GA Algorithm

1. The algorithm used for the Tabu GA is given below:

2. Initialise the population,

3. Initialise the Tabu List,

4. Perform basic G.A operation after checking the Tabu list, and

5. If the basic GA operators are disturbing the Tabu list, leave the particular position without modification.

6. Repeat the operations for 100 generations.

5. Convergence of Tabu Genetic Algorithm

This section is used to analyse the convergence of Tabu Genetic Algorithm (Tabu GA). The convergence is compared with the odd and even point cross overused in the previous section. The convergence analysis is carried using the following two considerations as mentioned in the last section.

1. Average fitness value, and

2. Variation in the average fitness values from one generation to other.

Table 4. Percentage of maximum and minimum difference

DataSet	Minimum	Maximum
ADI	0.082	9.991
CISI	0.225	6.189
CRAN	0.145	7.79

5.1 Average Fitness Value

The average fitness values of the individuals present in each generation are recorded, and it is tabulated. The main intention of this work is to analyse the convergence of Tabu Genetic Algorithm. The convergence of Tabu GA is compared with odd and even point crossover's convergence. Table 5 and 6 are compared, and there is a huge difference among the minimum and maximum value. The result shows smooth convergence for Tabu GA against the odd and even point crossover. Again, we want to confirm the results by using the statistical methods.

The difference in the average fitness value for each successive generation is compared using student – t-test. The hypothesis used for the testing is given below:

H0: There is no difference between the mean value for two population: $\mu_1 = \mu_2$

H1: There is a significant difference among the average value for the two populations: $\mu_1 \neq \mu_2$.

6. Results and Discussions

The experiments are conducted over the same datasets used in section 2. The characteristics of the datasets and

Table 5. Tabu tenure period

Sl. No.	Name of components	Tabu tenure period
1.	Fusion Function	2
2.	Retrieval Strategies	1
3.	Weights	0

Table 6. Tabu GA based selection result

Collection	Precision	Percentage of Improvement
ADI	0.3812	6.809
CISI	0.1923	5.66
CRAN	0.1437	4.74

Table 7. Percentage of maximum and minimum difference

DataSet	Minimum	Maximum
ADI	0	0.8291
CISI	0	3.335
CRAN	0	0.856

the components are described in section 2. The results obtained for the Tabu GA is given in the Table 7. The Table 7 shows a slight improvement in the performance. The improvement in the performance is due to the exploitation of search space. The main intention of this work is to test the convergence of the Tabu GA. The convergence analysis has been carried out in the next section.

The average precision value for all the data sets is recorded and plotted. The plotted graph is shown in Figures 1-3. The graph shows the steep fluc-
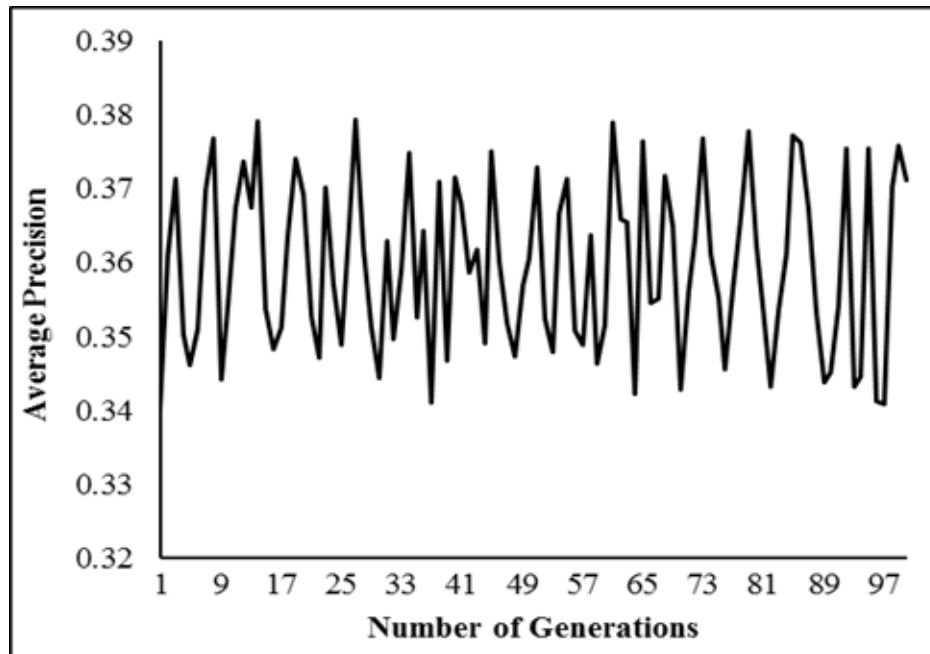
Figure 1. Average precision for 100 generations over ADI test collection.
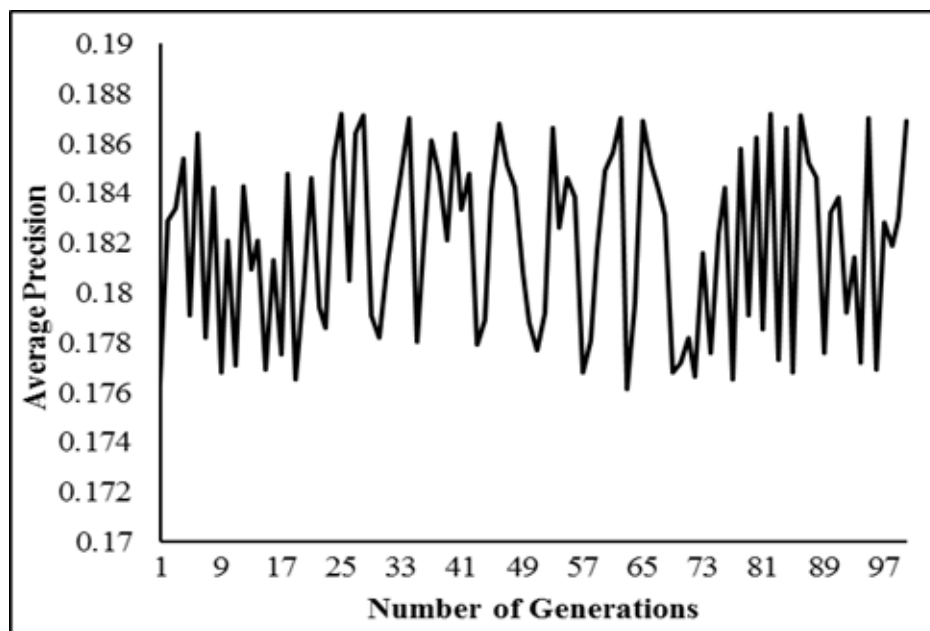
Figure 2. Average precision for 100 generations over CISI.

tuation from one generation to other generation. We have not got the same results for any two successive generations. It is an indicator of the vigorous exploration and slow convergence. The absolute value of the difference between two successive generations is calculated, tabulated and plotted as the bar chart. The chart, which shows
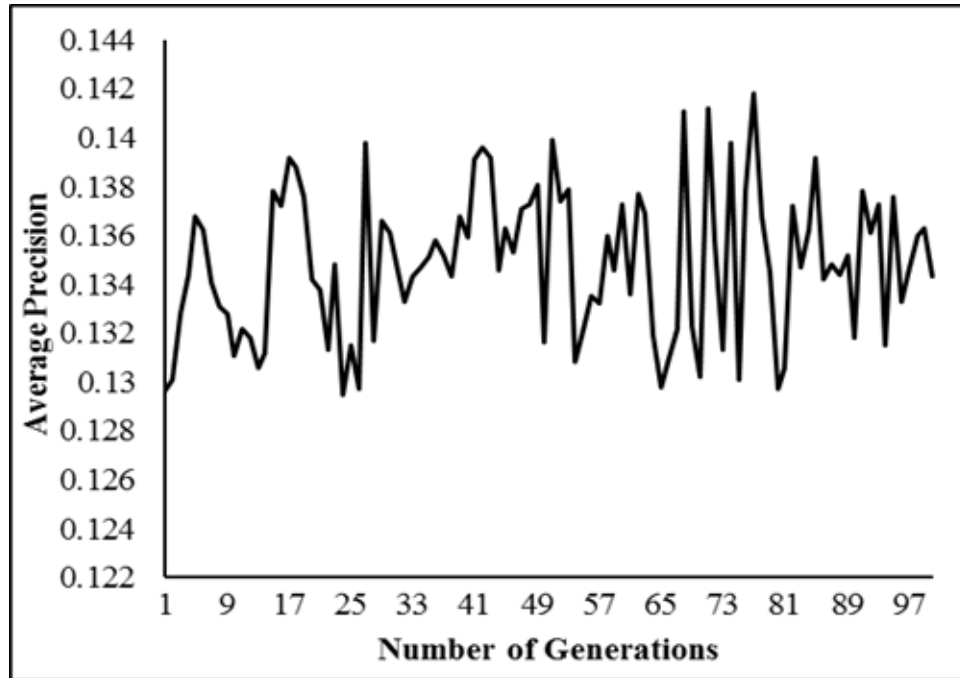
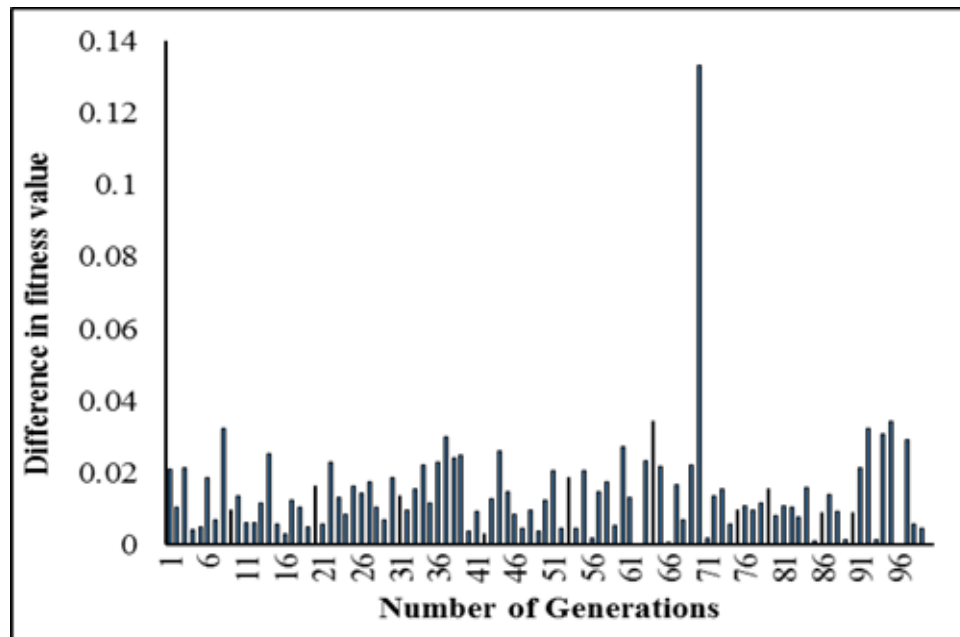**Figure 3.** Average precision for 100 generations over CRAN.

Figure 4. Difference in fitness value between successive generations over ADI test collection.

the difference, is given in Figures 4-6. Figures 4-6 show the rapid fluctuation. The step size between the bars is also high. The difference between successive generations is computed as the percentage of deviation. The percentage of deviation is tabulated, and we give the minimum and the maximum value in Table 4. The Figures 1-6 and
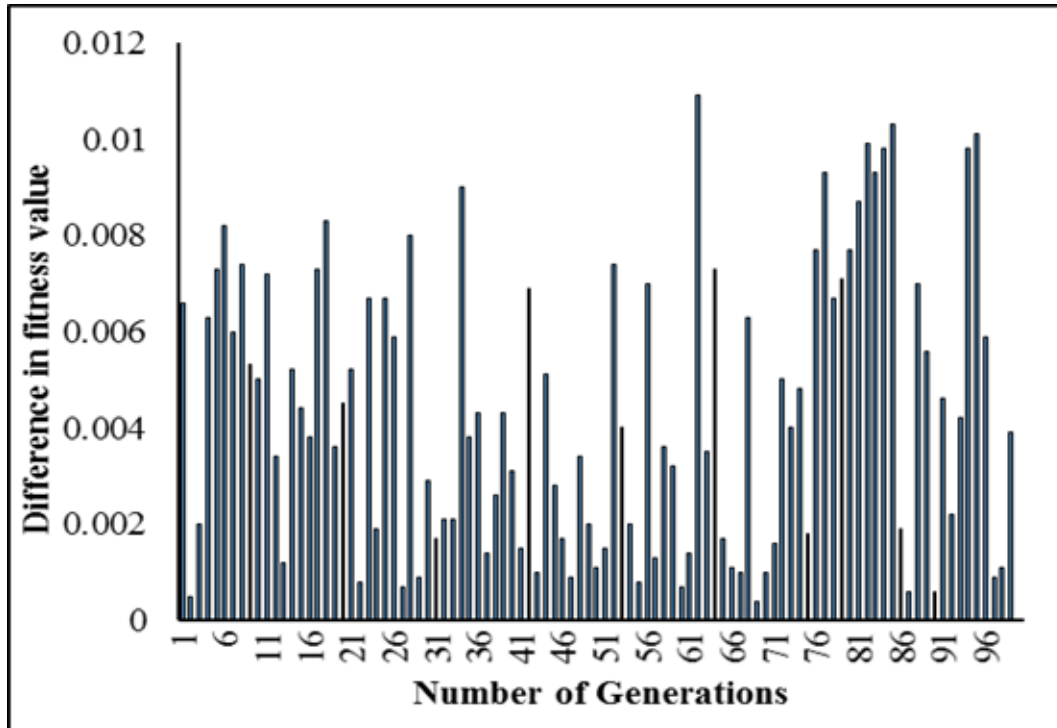
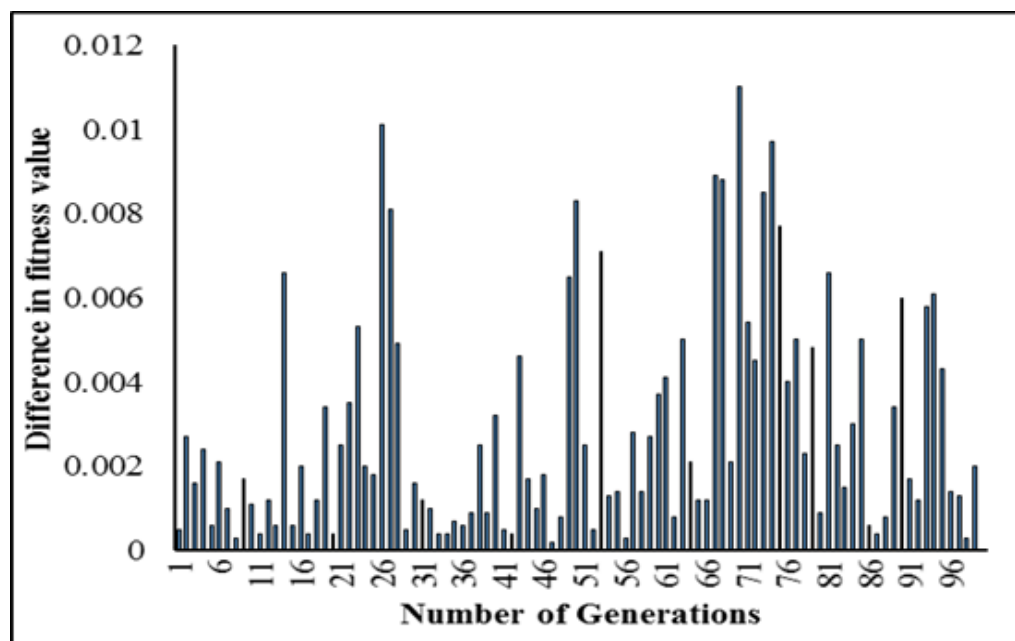**Figure 5.** Difference in fitness value between successive generations over CISI test collection.

Figure 6. Difference in fitness value between successive generations over CRAN test collection.

Table 4 show the slow convergence of the odd and even point cross over. The convergence is a relative property, and we want to test whether there is improve mention the convergence by merging the best local optima, i.e. Tabu with the GA. The convergence of Tabu GA is analyzed in section 5. The Figures 7-9 show the average fitness value
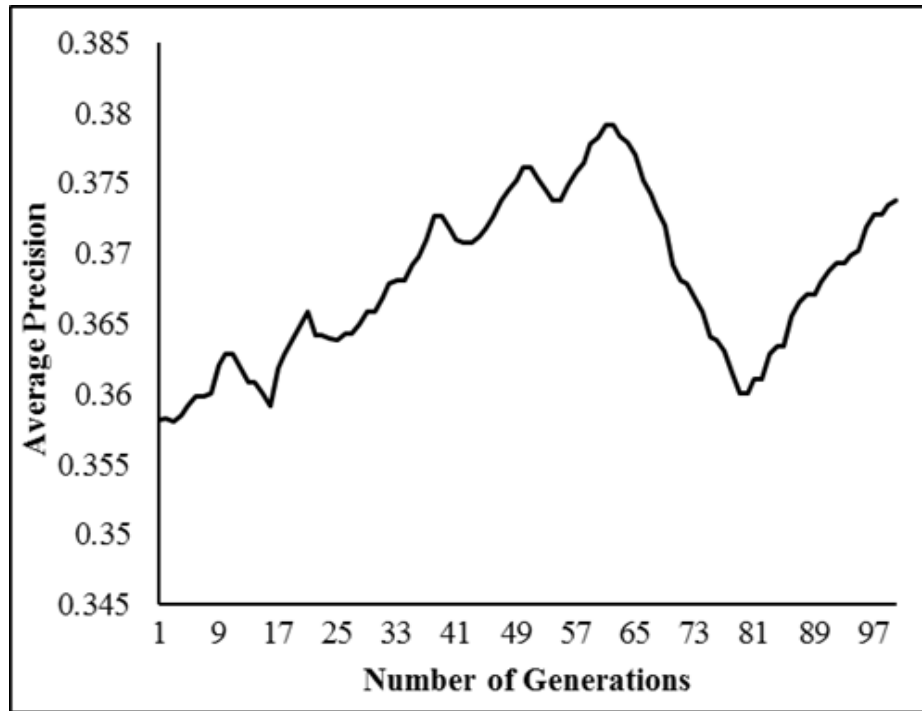

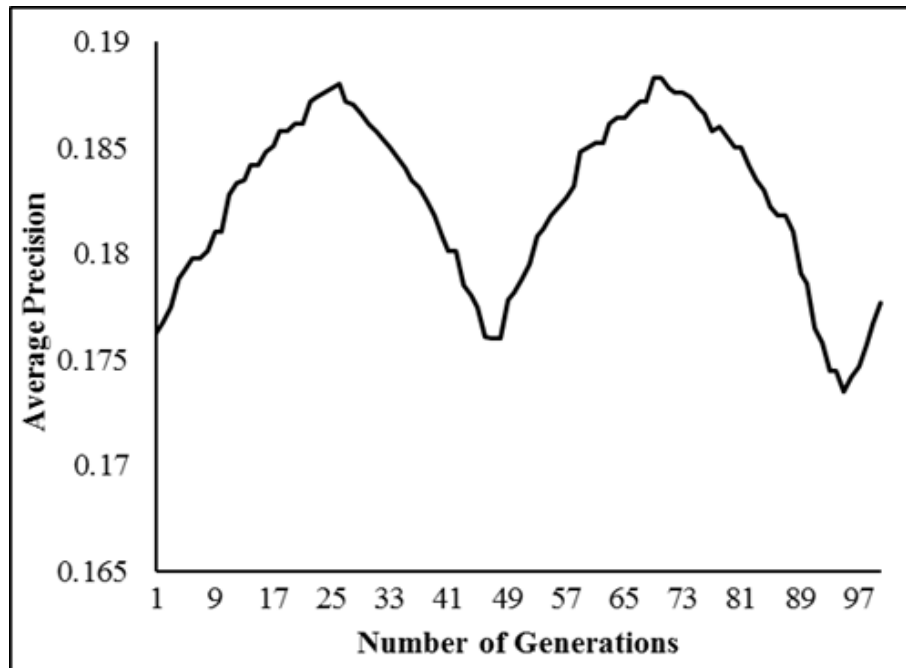**Figure 7.**   Average precision for 100 generations over ADI.

Figure 8.   Average precision for 100 generations over CISI.

for the three data set's for 100 generation. The transition from one generation to the other is smooth. We have calculated the difference between the fitness values for each successive generation. The difference in the fit-
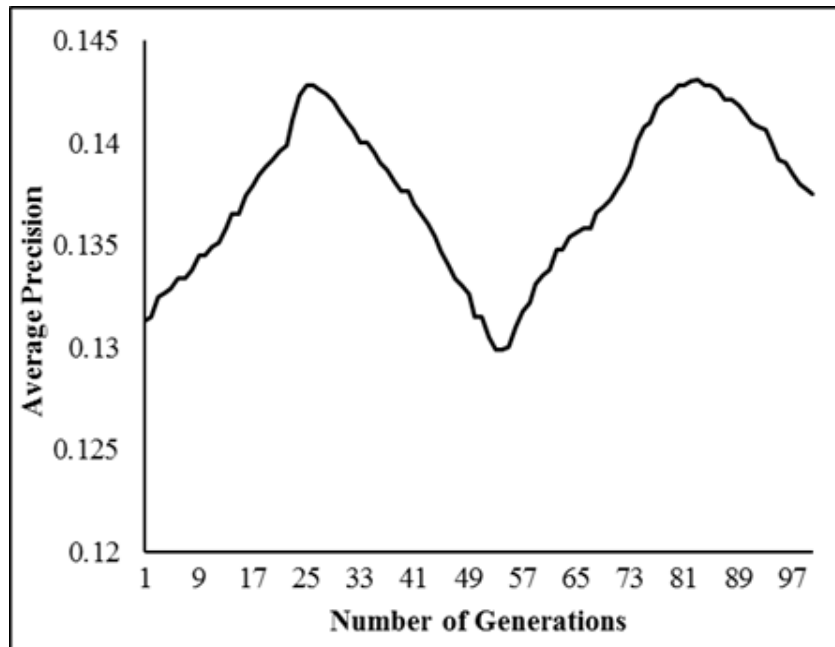

Figure 9.   Average precision for 100 generations over CRAN.
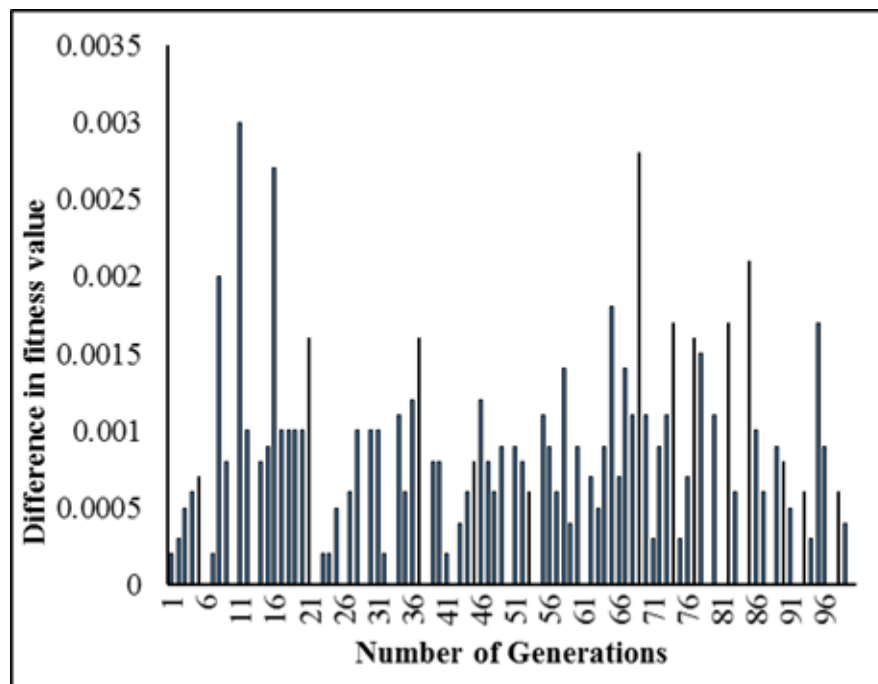

Figure 10.   Difference in fitness value between successive generations over ADI.

ness value is plotted, and it is shown in the Figures10-12. Figures 10-12 show the fluctuation in fitness value from generations one to other. The step size is small, and the percentage of deviation from one to other is calculated and tabled in the following Table 7. The computed t value is given in the Table 8. The NULL hypothesis is success-
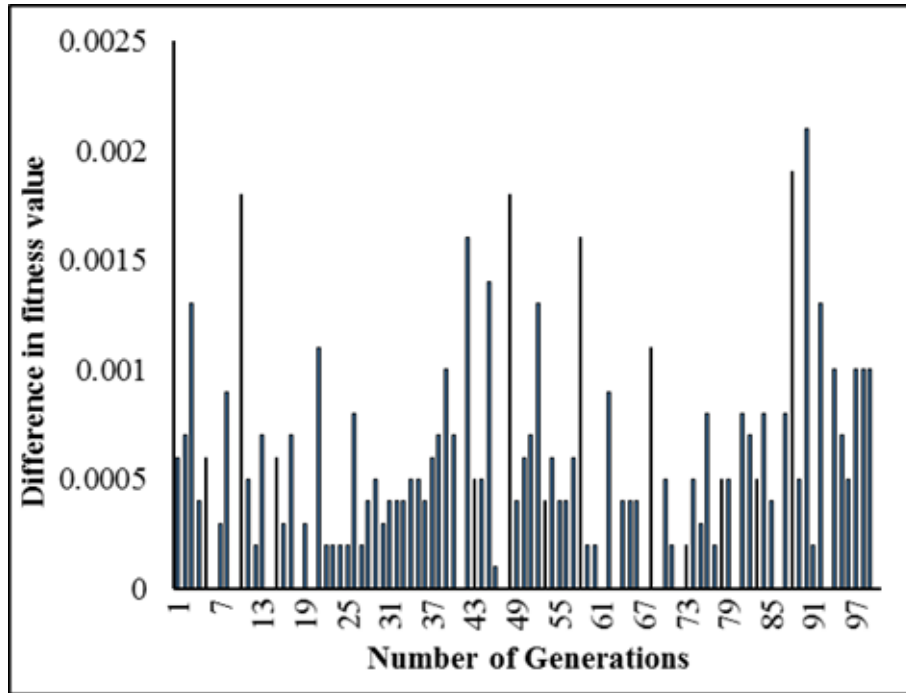

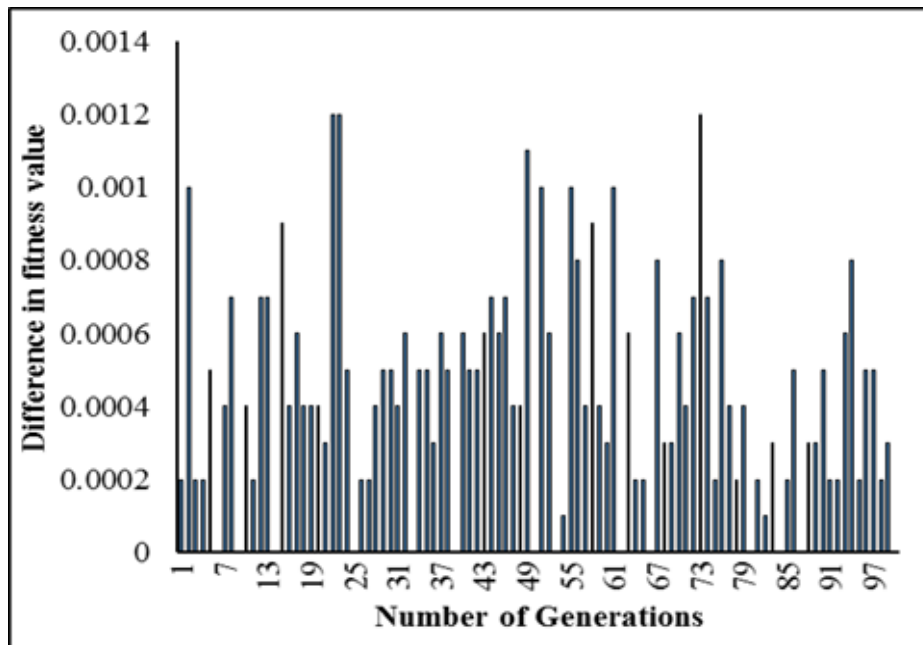**Figure 11.** Difference in fitness value between successive generations over CISI.

Figure 12. Difference in fitness value between successive generations over CRAN.
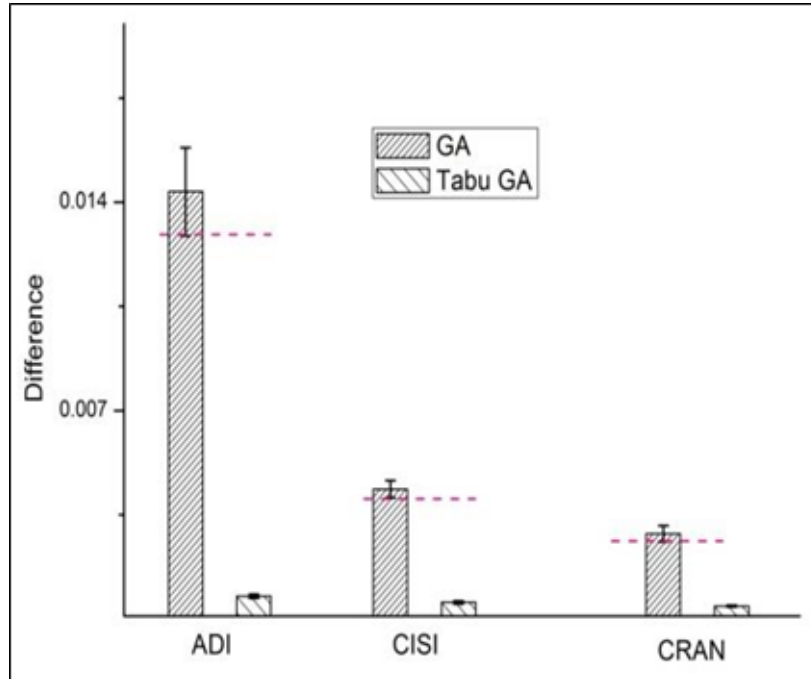
Figure 13. Odd and even point cross over and Tabu GA comparison.

fully rejected at the 0.001% confidence level. The Figures 7-9 and 10-12 and the computed t value confirms the smooth and quick convergence Tabu GA. The Statistical parameters are plotted in the Figure 13. The figures indicate a huge gap between the two populations and confirm the smooth convergence.

Table 8. Computed t value

DataSet	T Value
ADI	9.169
CISI	12.860
CRAN	9.205

7. Conclusions

The exploitation of search space has been carried out using Tabu GA. At the end of average fitness value of the individual the experiments unconscionable fluctuation from one generation to other generation. We have not to find the same effects for some two successive generations.

Durability is an indicator of the energetic exploration and slow convergence. The convergence of Tabu GA has been tested, and it has been proved, the Tabu GA has smooth convergence over the conventional GA. The experiments establish a relationship between GA and Tabu GA. If the Tabu tenure period is '0', then Tabu GA becomes conventional GA. If the tenure period is high, then the Tabu GA becomes Tabu Search. Hence, there should be a trade-off among Tabu tenure value. The impact of Tabu tenure period on Tabu GA is not studied thoroughly and impact of Tabu tenure period over the odd and event point cross-over based Tabu GA, we intend to carry out this work in the near future. We want to merge and test other possible local search algorithm with the GA. By doing so, we want to produce new type of hybrid search tool.

8. References

1. Salton G, McGill MJ. Introduction to modern Information Retrieval. McGraw-Hill; 1983. p. 1–448.
2. Yates RB, Neto BR. Modern Information Retrieval. Addison-Wesley; 1999. p. 1–103.

PMid: 10188590.

3. Korfhage RR. Information storage and Retrieval. Willey computer Publishing; 1997. p. 1–349.

4. ZobelJ, Moffat A. Exploring the similarity space, ACM SIGIR Forum. 1998; 32(1):18–34.

5. Lee JH. Combining Multiple Evidence from Different Properties of weighting schemes. Date accessed: 13/07/1995. https://dl.acm.org/citation.cfm?id=215358.

6. Lee JH. Combining Multiple Evidence from different relevance feedback network, Database Systems for Advanced Applications. 1997; 97:421–30.

7. Billhart H. Learning retrieval expert combinations with genetic algorithms, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2003; 11(3):87–113. https://doi.org/10.1142/S0218488503001965.

8. Fisher HL, Elchesen DR. Effectiveness of combining title words and index terms in machine retrieval searches; Journal of Nature.1972; 109–10.

9. Lee JH. Analyses of Multiple Evidence Combination. Proceeding of the ACM SIGIR Conference on Research and Development in Information Retrieval; 1997. 267–76. https://doi.org/10.1145/258525.258587.

10. Fox EA, Shaw JA. Combination of multiple searches, Proceeding of the Second Text Retrieval Conference (TREC-2). 1994; 500-215:243–52.

11. Fox EA, Shaw JA. Combination of Multiple Searches. Proceeding of the Third Text Retrieval Conference (TREC-3); 1995. p.105–08.

12. Gold Berge DE. Genetic Algorithms in Search, Optimization, and Machine learning. Addison-Wesley; 1989. p. 1–432.

13. Information Retrieval: A Survey. Date accessed: 30/11/2000. https://www.csee.umbc.edu/csee/research/cadip/readings/IR.report.120600.book.pdf.

14. Chor B, Goldreich O, Kushilevitz E, Sudan M. Private Information Retrieval, Journal of the ACM. 1998; 45(6):965–82. https://doi.org/10.1145/293347.293350.

15. Data Fusion. Date accessed: 15/01/2009. https://dl.acm.org/citation.cfm?id=1456651.

16. Tabu Search. Date accessed: 2011. https://wiki.eecs.yorku.ca/course_archive/201112/F/4403/_media/tabu_search.pdf.

17. Tabu Search Fundamentals and Uses. Date accessed: 1995. https://www.researchgate.net/publication/249776329_Tabu_Search_Fundamentals_and_Uses.

18. Senaratna NI. Genetic Algorithm: The crossover-Mutation Debate. A literature survey (CSS3137-B) submitted in partial fulfillment of the requirements for the Degree of Bachelor of Computer Science (Special) of the University of Colombo; 2005. p. 1–26.