

Analysing Genetic Algorithm Techniques on Data Mining for Constructing Effectual Business Intelligent System (EBIS)

*Antony S Alexander *C. Jothivenkateswaran

Abstract : *Genetic Algorithms*, provides continuous and discrete functions to optimization and search to be carried across various data sets. *Genetic Programming*, are evolved to program computer systems as an approximate problem solving methods. Genetic algorithms is take an important role, to find complex spacing handling and in many other fields namely financial, marketing, engineering, automated systems, AI etc. This paper deals with the specific utilization of genetic algorithm across data mining domain. The data mining techniques like Prediction, clustering and micro clustering are effectively done by implementing the successful methods of mutation and crossover of the Genetic Algorithm. It is based on the genetal process, henceforth it can be used to design and brilliant business intelligent system which works on the words population principles.

Keywords : Genetic Algorithm(GA), Artificial Intelligence(AI), Business Information(BI), Mutation, Crossover, Micro Clustering (MC)

1. INTRODUCTION

Many researches have been done on various domains with the data sets for building an Optimal Business Intelligent Systems. The resultant system after interpreting with the huge data sets of the research will lead to a Business Intelligent System (BI) with contemporary technologies. Now a day's data oriented study has become the core area of research. Data has become the intellectual property to be preserved and pirated. With the available data sets, so many types of analysis are done for building various computational systems which has become core for various business operations as public domain. Thus intelligent data based predictions has become vital assert for successful business process. The application which foresees the business process for certain affirmative decision making will use these findings for their effective business process.

Business Risk is fundamental part of financial services. Managing the risk is a complex task for any financial business, and gradually more important in a world where economic events and financial systems are associated. Global financial society regulators have worked on risk management with various methods for long-term success (*J. T. Alander*)[1]. The risk management are mitigated by foreseeing the risk and the systems are designed by integrating various techniques which includes Artificial Neural Network (ANN), Support Vector Machine (SVM), the Regressive Analysis (RA), statically analyzing of data sets. The Genetic Algorithm (GA) methods are used in this is work for finding out the effective simple optimized system using the crossover and mutational Techniques. The performance of these techniques has become one of the most important initiative to manage the microfinance industry across the public domain(*Silltow J*)[13].

This study prepares to presents the Genetic Algorithmic technique and implemented in on a system, the findings are implemented to assess the efficiency of a financial structure. This competent system largely supports the fund flow of Micro investments made at the normal Clusters of Localized system[4][5].

2. REVIEW OF LITERATURE

Newman (2013) also discusses that a lot of good things can come from data mining techniques. He has places of interest the top benefits of data mining as to invest more money as a result of profits and investments; patter into new markets strategies; the principle of share and share alike involves sharing of information that may be useful to related organizations; learning from the past; help in challenger analysis where data mining helps companies to get info that they can use effectively to put out from competition[6].

According to *Silltow (2006)*, data mining automates the discovery of relevant patterns in a database, using defined approach and algorithms to look into present and past data that can be analyzed to predict future trend. Because data mining tools predict future trend and behavior by reading across the databases for hidden patterns, these patterns allow organizations to make positive, knowledge-oriented decisions and answer questions that were too time-consuming to resolve a solution. *Silltow (2006)* lists the hub of data mining techniques as: association, classification, prediction, link analysis, sequence analysis, clustering, sequential patterns, decision trees, combinations, long-term (memory) processing.

Although research by *FSD (2008)* and show that most microfinance institutions uses the credit scoring model for data mining, where the data classification is used to analyze credit risk, it is applied on the dataset of the previous customers available at the financial institution to distinguish them as good or aberrant customers, to find if there exists a association between the characteristics and reasons for delinquency of loans and consequently choose an accurate classifier to implement on the new applicants. This paper focuses on comparing different Genetic algorithmic techniques used with data mining methods when applied to such data[2][3].

3. METHODOLOGY

The data mining techniques like clustering and classification is done for developing Business Intelligent (BI) systems. In this paper a BI system has been derived by implementing micro financial data. The effectual crossover and mutation methods have been used to cluster the micro elements of the data set.

Genetic Algorithm (GA) is a global optimization algorithm derived from evolution and natural selection. GA provides best solution for various problems even though GA cannot provide optimal solution always [7][8][9].

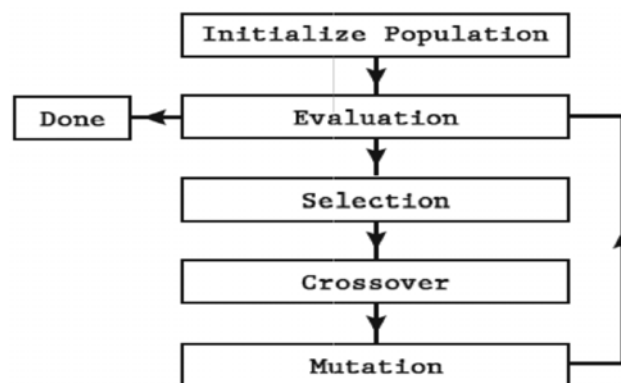


Fig. 1. Genetic Algorithm.

1. Three major operations of GA are;
2. assessment of individual fitness,
3. forming of a gene pool (intermediate population) using selection mechanism,
4. Recombination is done by using crossover and mutation operators.

The basic thought of Genetic algorithm:

Step 1 : Produce initial population.

Step 2 : Produce next generation by crossing over and mutation methods.

Step 3 : New population from step 2

Step 4 : Produce the next population by repeating the Step 2 and Step 3 until obtaining the individual which satisfies.

Pseudo Code Genetic Algorithm (GA)

Begin

N = 0;

initialize P(N);

evaluate P(N);

While (Not *termination-condition*) do

begin

N = N + 1;

select P(N) from P(N-1);

recombine P(N);

evaluate P(N);

end

end

Next procedure shows the structure of a basic GA, where P(N) denotes the population at generation N.

Genetic Algorithm consists of two operational techniques called Crossover and mutation explained below. Its Performance depends on the techniques. Hence applying the equations the algorithm of our GA for Micro financial system using Crossover and Mutation can be described as follows.

A. Crossover

Crossover is a genetal operation that combines two individual characters, to produce a new character. The technique behind crossover method is to derive a new child may be better than both of the parents. It takes the best characters from each of the parent characters. Crossover occurs through evolution according to user – definable crossover probability. Crossover selects individual character from parent characters and creates a new child. In Genetic Algorithm randomly chooses two individuals from population and changing the bits (character) of the same section of the two individuals.

In mathematical equation Crossover operation linearly combines two parent bits vectors to produce two new Childs according to the equations

$$N_1 = a * P_1 + (1 + a)* P_2$$

$$N_1 = (1 + a)* P_1 + a * P_2$$

Where ‘a’ is random weight factor chose before each crossover operation.

Consider two parent (each of 4 float bits) selected for crossover

P1	0.3	1.4	0.2	7.4
P2	0.5	4.5	0.1	5.6

Applying the above two equation and assume the weighting factor $a = 0.7$, we get two resulting new bit.

The possible set of bit after arithmetic crossover would be

P1	0.36	2.33	0.17	6.87
P1	0.402	2.981	0.149	5.842

B. Mutation

After a crossover, mutation takes place. Mutation is a genetal operation used to maintain genetic diversity form one generation of a population to the next generation. Mutation occurs during development according to a user-definable mutation probability, usually set to fairly low value say 0.02 as a good first choice[10].

Mutation alters one or more attribute values in the genetal characters from its initial state. This can result in completely new attribute value being added to the gene pool. With the new character values, the genetic algorithm may be able to reach out to a better solution than was previously possible[11][12].

Mutation is an vital part of the genetic clustering, it helps to averts the population against the investment done on macro financial systems. In example this operator simply inverts the value of the chosen bit. Such as 0 goes to 1 and 1 goes to 0.

This operator can only be used for binary genes. Consider the original off-springs selected for mutation.

P1 1 1 0 1 1 1 1 0 0 0 0 1 1 1 1 0

P2 1 1 0 1 1 0 0 1 0 0 1 1 0 1 1 0

Invent the value of the chosen gene as 0 to 1 and 1 to 0.

The mutated bits stream produced are:

P1 1 1 0 0 1 1 1 0 0 0 0 1 1 1 1 0

P2 1 1 0 1 1 0 1 1 0 0 1 1 0 1 0 0

4. RESULT AND DISCUSSION

This paper presents the findings of the study using Genetic Algorithm for data mining techniques, to be implemented on a financial system, in search of to find out the optimal GA technique to be analyzed with the performance of microfinance data set are recommended and commonly used on different literate review to select the best accuracy.

The data was used given in to MATLAB and the experiments are

Table 1. Experiential Table in MATLAB

<i>Genetic Algorithm</i>	<i>Training Data (%)</i>	
	<i>Accurate</i>	<i>Inaccurate</i>
Crossover	69.87	30.13
Mutation	70.15	29.85

The Table-1 shows the Sample data for Training data set was tested for Accurate and inaccurate results. For the given data under this study GA 10 fold cross validation has shown better accuracy and the rule generated by this model was used for interpretation and for easy understanding the of the experiments[13][14].

Table 2. Experiential Table in MATLAB.

<i>Genetic Algorithm</i>	<i>Test Sample (%)</i>	
	<i>Accurate</i>	<i>Inaccurate</i>
Crossover	67.89	32.11
Mutation	73.15	26.85

The Table-2 shows the micro financial data set being tested with huge data set. In this tested results accurate and inaccurate results are analyzed in Table-2, to evaluate the performance of the Genetic Algorithm techniques. The results are shown in below graphical representation.

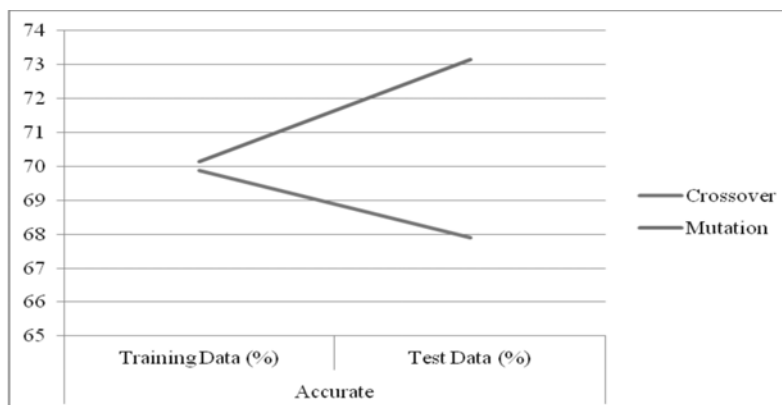


Fig. 2. Valid Respresentationa.

The above Figure-2 shows the accuracy rate of Crossover and mutation techniques of Genetic Algorithm.

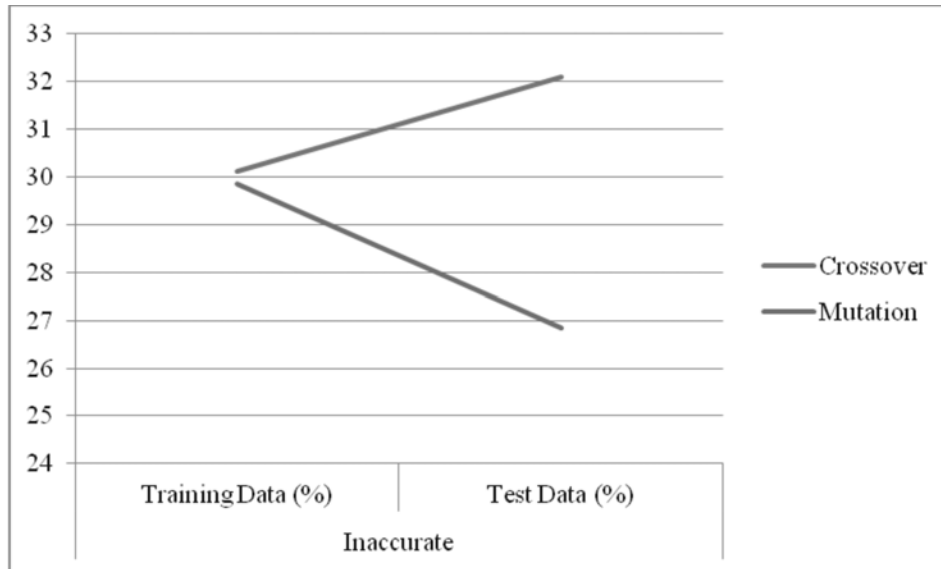


Fig. 3. InValid Respresentationa.

The above Figure-3 shows the Error rate of Crossover and mutation techniques of Genetic Algorithm

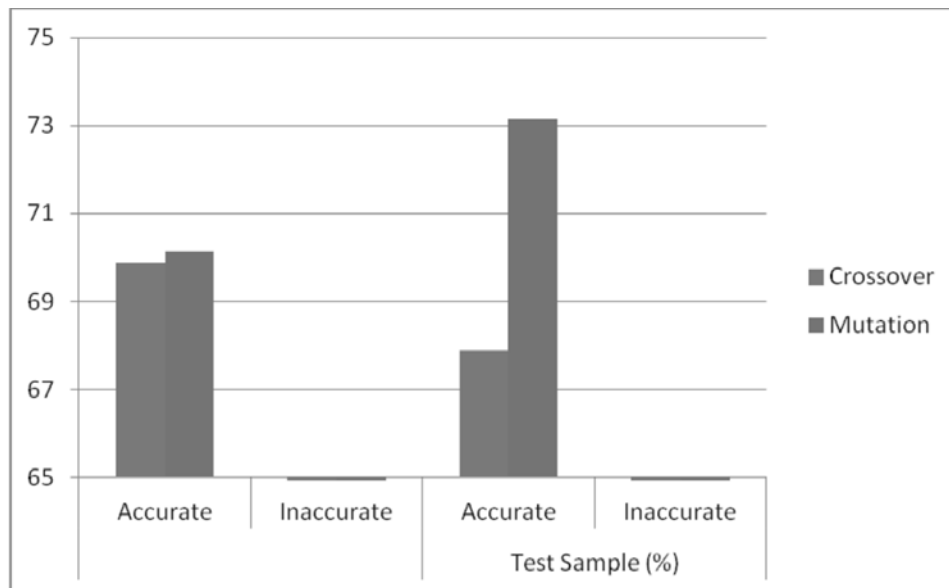


Fig. 4. InValid Respresentationa.

The above Figure-4 shows the comparative representation of the Crossover and mutation methods of Genetic Algorithm against the accuracy with Error rate.

5. CONCUSSION

The findings of the study were guided to study the objective in determining the impact of the data mining techniques on performance of Microfinance system data set. The experiential results of Table-1 and Table-2 clearly shows that Genetic Algorithm methodologies are efficient for the implementing the data mining techniques. The experiment model building was done using crossover and mutation with the dataset and attributes. The experiment resulted in an accuracy of 73.15% for Mutation and 67.89% for crossover, gives, the best accuracy rates both in test sample and micro financial data set which indicates clearly that mutation is a better classification technique when implemented for the Data mining technique.

6. REFERENCE

1. J. T. Alander. An indexed bibliography of genetic algorithms and fuzzy logic. Technical Report 94-1-FUZZY, Department of Information Technology and Industrial Management, University of Vaasa, Finland, November 1996. Available at http://www.uwasa.fi/_jal.
2. Liang jiye. (1999). The research of common problems in genetic algorithm application. *Research of Computer Application*, 1999(7): 20-21.
3. FSD (2008). The Potential for Credit Scoring for SME lending in Kenya, Vol.1
4. Liu guohua, Bao hong & Li wenchao. (2001). The genetic algorithm programming in MATLAB. *Research of Computer Application*, 2001(8):80-82.
5. Liu yong, Kang lishan & Chen yuping. (1997). *Nonnumerical parallel algorithm(second volumn)—genetic algorithm*. Beijing:science press, 1997.
6. Newman D., (2013). Introduction to Data Mining. California: Irvine
7. Xi yugeng, Chai tianyou & Yun weimin. (1996). Summarization of genetic algorithm. *Control Theory and Application*, 1996(6):697-708.
8. Yin ming, Zhang xinghua & Dai xianzhong. (2000). The programming of genetic algorithm in MATLAB. *Application of Electronic Technology*, 2000(1):9-11.
9. Zhou ming & Sun shudong. (1999). *Theory and Application of Genetic Algorithm*. BeiJing: national independence industry press, 1999.
10. Cheng-Lung Huang a, Mu-Chen Chen b, Chieh-Jen Wang, “ Credit scoring with a data mining approach based on support vector machines”, *Expert Systems with Applications* 33 , 847–856.
11. Haron.O.Moti,Justo simiyu Masinde,Nebat Galo Mugenda ,Mary Nelima Sindani, ”Effectiveness Of Credit Management System on Loan Performance: Empirical Evidence From Microfinance sector in Kenya”, *International Journal Of Business ,Humanities and technology* , Vol.2,No 6 ,October-2012.
12. Noman, Syeda Shabih Fatima, and Shaheera Rachel Shahlal. “An Empirical Analysis of Financial Performance of Conventional Banking Sector in Islamic Republic of Pakistan.” *Global Journal of Management And Business Research* 16.2 (2016).
13. Silltow J., (2006). Data mining tools and techniques. United Kingdom
14. <http://www.neurodimension.com/genetic/mutation.html>